

版权注意事项：1、书籍版权归著者和出版社所有；
2、本PDF仅用于个人获取知识，进行私底下知识交流；
3、PDF获得者不得在互联网以任何目的进行传播；
如有需要，请尽量购买正版实体书！支持书籍作者！！



“十二五”国家重点图书出版规划项目

大数据技术与应用

丛书策划

上海大数据产业技术创新战略联盟（上海产业技术研究院）

上海市数据科学重点实验室（复旦大学）

丛书主编

朱扬勇 吴俊伟

Big Data

Technology and Application Series

李光亚 张鹏翥 孙景乐 等
编著

智慧城市 大数据



上海科学技术出版社



大数据技术与应用

智慧城市大数据

李光亚 张鹏翥 孙景乐 等
编著

上海科学技术出版社

本书出版由上海科技专著出版资金资助

图书在版编目(CIP)数据

智慧城市大数据 / 李光亚等编著. —上海: 上海
科学技术出版社, 2015. 1(2015. 1 重印)

(大数据技术与应用)

ISBN 978 - 7 - 5478 - 2290 - 6

I. ①智… II. ①李… III. ①现代化城市—城市建设
—研究 IV. ①C912. 81

中国版本图书馆 CIP 数据核字(2014)第 145252 号

智慧城市大数据

李光亚 张鹏翥 孙景乐 等 编著

上海世纪出版股份有限公司 出版
上海科学技术出版社

(上海钦州南路 71 号 邮政编码 200235)

上海世纪出版股份有限公司发行中心发行

200001 上海福建中路 193 号 www. ewen. co

苏州望电印刷有限公司印刷

开本 787×1092 1/16 印张 14

字数: 310 千字

2015 年 1 月第 1 版 2015 年 1 月第 2 次印刷

ISBN 978 - 7 - 5478 - 2290 - 6/TP · 28

定价: 52.00 元

本书如有缺页、错装或损坏等严重质量问题, 请向工厂联系调换

内 容 提 要



本书阐述了智慧城市的技术体系总体框架,围绕该框架,对智慧城市中的大数据信息资源进行了分类,进而介绍大数据在智慧城市应用中的关键技术。除对可能的共性应用场景进行论述外,本书还针对民生领域、市场监管、政府服务、基础设施等四大重点类别进行详细分析,包括大数据在这些智慧城市应用领域中的数据资源、应用场景以及发展趋势。

本书可供对大数据技术在智慧城市中应用感兴趣的相关技术、业务、咨询规划人员阅读,同时也可为企业家、投资家和政府工作人员等提供参考。

大数据技术与应用

学术顾问

中国工程院院士 邬江兴

中国科学院院士 梅 宏

中国科学院院士 金 力

教授,博士生导师 温孚江

教授,博士生导师 王晓阳

教授,博士生导师 管海兵

教授,博士生导师 顾君忠

教授,博士生导师 乐嘉锦

研究员 史一兵

大数据技术与应用

编撰委员会

主 任

朱扬勇 吴俊伟

委 员

(以姓氏笔画为序)

于广军 朱扬勇 刘振宇 孙景乐 李光亚 李光耀 杨 丽
杨佳泓 吴俊伟 何 承 张鹏翥 陈 云 武 星 黄林鹏
童维勤 蔡立志

本书编委会



当前,城市发展面临种种挑战,例如人口膨胀、环境恶化、公共	主 编	频发、交通
拥堵、资源浪费等,人们的生活也因此受到严重影响。随着	李光亚	化文明进
步的推进,人们日益扩大的需求规模与城市日益有限的发展	张鹏翥	续增大,未
来人们的生活质量不容乐观。	孙景乐	
智慧的地域、智慧的城市是当今世界的热点,也是发展的趋势。		
2013年,我国	编 委	互联网结合
已有30多个城市提出或正在建设智慧城市。智慧城市是指在物联网	陈 诚	行管理到
的基础上,通过各种信息基础设施实现信息资源的共享和整合,提	王兆进	理能力的
高政府公共服务水平,促进经济社会可持续发展。	周志中	建设巨大到
大数据源于结构化和非结构化数据。随着信息技术的	崔有祥	企业经营
发展,使得数据量级不断提高。从概念上讲,大数据是指无法	李剑华	企业数据
用传统的数据工具在合理时间内采集、管理、处理和建立	童 庆	数据的数
海量的资讯。美国已经将大数据提升为	孔维亚	数据的数
国家卫生信息共享技术及应用工程技术研究中心	凌 力	析和利用。
万达信息股份有限公司	肖筱华	管理城市,
上海交通大学		这是当前
上海交通大学		非常重要的城市发
万达信息股份有限公司		展定位。智慧城市建设,迫切需要使用先进的技术,包括数据
万达信息股份有限公司		挖掘和功
上海市市民信箱管理中心		能强大的运算系
国家卫生信息共享技术及应用工程技术研究中心		统,来整合分析跨地域、跨行业、跨部门的海量数据,将特定的知识应
		用于特定的行业和特定的解决方案中,支持整个经济社会发展的决策和相关行动。
		从智慧城市体系结构来看,相对于之前的数字城市概念,最大的区别在于对感知

前言

当前,城市发展面临种种挑战,例如人口膨胀、环境恶化、公共卫生事件频发、交通拥堵、资源浪费等,人们的生活也因此受到种种不利的影响。而且,随着城市化文明进程的推进,人们日益扩大的需求规模与城市日益有限的供给之间的矛盾会继续增大,未来人们的生活质量不容乐观。

智慧的地球、智慧的城市是当今世界的热点,也是发展的趋势。截至 2013 年,我国已经有 311 个城市提出或正在建设智慧城市。智慧城市是指在物联网和互联网结合的基础上,通过各种智能化的应用,提升城市基础设施的运作效率,提升城市运行管理和公共服务水平,让人们的生活更美好。

大数据源于结构化和非结构化数据的猛增,包括互联网、物联网、计算处理能力的发展,使得数据量级不断提高。从概念上讲,大数据指的是所涉及的资料量规模巨大到无法采用目前主流软件工具在合理时间内撷取、管理、处理和整理,成为帮助企业经营决策的资讯。美国已经将大数据提升为国家战略,中国还没有明确提出,但已经把大数据上升为与国防一样的高度,多部委还联合发布了鼓励措施。我国政府对大数据的敏感度快速提高,并正在采取措施。

从智慧城市的根本目标来看,所谓的“智慧”必然来自对大数据的充分分析和利用。因此,如何对数据进行采集、分析和利用,促进人类智慧地运用、感知、服务与管理城市,这是当前面临的挑战。利用收集的数据,全面、深入、系统地解决城市面临的问题,这是非常重要的城市发展定位。智慧城市建设,迫切需要使用先进的技术,包括数据挖掘和功能强大的运算系统,来整合分析跨地域、跨行业、跨部门的海量数据,将特定的知识应用于特定的行业和特定的解决方案中,支持整个经济社会发展的决策和相关行动。

从智慧城市的体系结构来看,相对于之前的数字城市概念,最大的区别在于对感知

前言

层获取的数据进行了大数据处理,从而获得了支撑和保障智慧城市顺利运营的多元信息,因此也可以简单认为智慧城市是数字城市的升级版。由城市数字化到城市智慧化,关键是要实现对数字信息的智慧处理,其核心是引入了大数据处理技术。在可以预见的未来,大数据将遍布智慧城市的方方面面,从政府决策与服务,到城市的产业布局和规划,到城市的运营和管理方式,以及到包含衣食住行在内的居民生活方式,都将在大数据支撑下走向“智慧化”,大数据将成为智慧城市的“智慧引擎”。

本书正是从智慧城市入手,阐述了智慧城市的概念和建设目标,分析了智慧城市在国内外的发展现状,由此思考了智慧城市所面临的问题和未来发展趋势。在此基础上,本书提出了智慧城市的技术体系总体框架,围绕该框架,对智慧城市中的大数据信息资源进行了分类阐述,并由此梳理了大数据在智慧城市应用中的关键技术。

大数据在智慧城市中的应用领域非常广泛,本书先是对可能的一些共性应用场景进行了描述,随后按照民生领域、市场监管、政府服务、基础设施等四大类别,详细分析了大数据在这些智慧城市应用领域中的数据资源、应用场景以及发展趋势。除了具体的技术和软件以外,产业分析、业务模式、标准规范、安全隐私等关键问题在智慧城市的建设过程中也是不可避免的,书中也对此进行了论述。

由于时间仓促和编委知识面所限,不足之处敬请读者指正!

作者

目 录

第1章 智慧城市概述

• 1.1 智慧城市的起源、概念和特征

1.1.1 智慧城市的起源

1.1.2 智慧城市的概念

1.1.3 智慧城市的特征

• 1.2 智慧城市的发展现状

1.2.1 智慧城市的发展

1.2.2 国外智慧城市现状

1.2.3 国内智慧城市现状

1.2.4 上海智慧城市现状

• 1.3 智慧城市的未来趋势

1.3.1 绿色宜居

1.3.2 平安有序

1.3.3 幸福和谐

1.3.4 集约高效

1.3.5 科学决策

参考文献

1

3

3

5

7

10

10

13

17

19

22

23

26

26

27

28

29

第2章 智慧城市的技术体系总体框架 31

• 2.1 智慧城市技术要求	32
2.1.1 创新性	32
2.1.2 实用性	32
2.1.3 智能性	33
2.1.4 扩展性	33
2.1.5 兼容性	34
2.1.6 安全性	34
2.1.7 规范性	35
• 2.2 智慧城市参考模型	35
2.2.1 智慧城市的概念模型	35
2.2.2 智慧城市的功能参考模型	37
• 2.3 智慧城市技术体系架构	44
2.3.1 总体技术架构	44
2.3.2 开放性体系	46
2.3.3 城市互联平台架构	48
参考文献	50

第3章 智慧城市中的大数据资源 51

• 3.1 智慧城市数据资源体系架构	53
3.1.1 数据资源体系	55
3.1.2 数据资源存储	57
3.1.3 基础数据资源库	58
3.1.4 共享数据资源库	60
3.1.5 特定数据资源库	62
3.1.6 公共服务数据资源库	64
3.1.7 管理数据资源库	67
3.1.8 政府决策数据资源库	67
• 3.2 智慧城市数据资源交换体系和目录体系	69
3.2.1 数据资源交换体系	70

3.2.2 数据资源目录体系

参考文献

73

76

第4章 大数据在智慧城市应用中的关键技术

77

• 4.1 数据生命周期管理

78

4.1.1 数据共享性：主数据管理

79

4.1.2 数据完整性：元数据管理

79

4.1.3 数据可靠性：数据清洗

80

4.1.4 数据有效性：数据存储策略

80

• 4.2 大数据采集

82

4.2.1 网络数据获取

82

4.2.2 媒体流获取

83

4.2.3 日志信息获取

83

4.2.4 传统应用数据获取

84

• 4.3 大数据存储

84

4.3.1 存储技术

86

4.3.2 硬件架构

89

4.3.3 CAP 原则及技术选择

91

4.3.4 大数据存储的未来

95

• 4.4 大数据处理

96

4.4.1 计算技术

96

4.4.2 数据挖掘技术

100

4.4.3 大数据可视化技术

102

• 4.5 智慧城市大数据决策分析

103

4.5.1 融合多领域信息大数据的知识模型

104

4.5.2 面向智慧城市决策支持的分布式大数据管理技术

104

4.5.3 辅助决策模型的智能组合技术

105

4.5.4 高性能多用户协同的辅助决策技术

105

参考文献

106

第5章 智慧城市中的大数据综合应用

• 5.1 网络大数据采集平台	109
5.1.1 互联网数据采集	110
5.1.2 物联网数据采集	112
5.1.3 个人数据采集	113
• 5.2 跨部门大数据管理平台	115
5.2.1 城市运营大数据	115
5.2.2 城市一体化公共信息平台	117
5.2.3 大数据平台服务	120
• 5.3 大数据开放服务平台	121
• 5.4 大数据决策支持平台	123
参考文献	124

第6章 大数据在智慧城市中的行业应用

• 6.1 智慧城市大数据应用的民生领域	126
6.1.1 领域大数据分析	126
6.1.2 典型应用场景	132
6.1.3 应用趋势分析	134
• 6.2 智慧城市大数据应用的市场监管领域	136
6.2.1 领域大数据分析	136
6.2.2 典型应用场景	138
6.2.3 应用趋势分析	146
• 6.3 智慧城市大数据应用的政府服务领域	147
6.3.1 领域大数据分析	147
6.3.2 典型应用场景	154
6.3.3 应用趋势分析	159
• 6.4 智慧城市大数据应用的基础设施领域	160
6.4.1 领域大数据分析	160

6.4.2	典型应用场景	161
6.4.3	应用趋势分析	162
	参考文献	162

第7章 大数据在智慧城市中的应用相关问题分析 165

• 7.1	产业分析	166
7.1.1	大数据应用对产业发展的意义	166
7.1.2	大数据在国内外的应用现状	169
7.1.3	市场规模和发展趋势	175
• 7.2	业务模式	180
7.2.1	商业模式的涵义	181
7.2.2	大数据业务的商业模式类型	181
7.2.3	大数据业务运营中的问题及对策	184
• 7.3	指标体系	185
7.3.1	概述	185
7.3.2	评价原则	187
7.3.3	评价指标	187
• 7.4	安全隐私	192
7.4.1	物联网的隐私维护和信息安全	193
7.4.2	云计算的隐私维护和信息安全	196
	参考文献	197

后记 199

索引 203

第1章

智慧城市概述

“智慧城市(Smart City)”一词最早于1984年由美国拉斯维加斯一家以“智慧城市”命名的产业技术协会组织提出。智慧城市是为解决城市病问题而提出的,目前没有公认的统一定义。简单来说,智慧城市就是让城市更聪明,本质上是让作为城市主体的人更聪明。智慧城市是一个复杂的社会大系统,通过自动传感、互联网、物联网、云计算等信息技术和自动化技术,使构成城市的基础设施、环境等要素与人相互之间实现互联、协同和智能,作为一个系统整体对民生、发展、环保、公共安全、城市服务、经济活动等需求做出智慧的响应。智慧城市的特征可从信息技术领域和社会领域两个方面阐述,其在信息技术领域的特征为:全面透彻的感知、物联网与互联网的完全整合、信息可靠传递、智能融合的应用;其在社会领域的特征为:人性化管理与服务、协同运作、以人为本的可持续创新。

美国是智慧城市发展的领跑者,早在2009年1月,IBM首席执行官彭明盛(Sam Palmisano)明确提出了“智慧地球”概念。美国政府将智慧地球上升到国家战略及政策层面,使很多陷入困境的企业看到了全新的希望。伴随着智慧地球概念的提出,IBM相继推出了各种“智慧”解决方案,包括智慧的电力、智慧的医疗、智慧的交通、智慧的供应链、智慧的银行业等,其中智慧城市是IBM智慧地球策略中的一个重要方面。新加坡2006年推出了为期10年的资讯通信产业发展蓝图——“智慧国2015”(或称“iN2015”计划),力图通过包括物联网在内的信息技术,达到成为一个由资讯通信所驱动的智慧国家与全球型都市的未来愿景。通过一系列项目和计划的实施,新加坡已在物联网建设方面走在了世界前列,在利用信息通信技术促进经济增长与社会进步方面也处于世界领先地位。瑞典的智慧城市建设在交通系统上得到了最大的体现。智慧改变交通,IBM和瑞典首都斯德哥尔摩的合作已经有了很好的示范。日本在2004年推出了基于物联网的国家信息化战略,称作“U-Japan”。该战略旨在推进日本信息通信技术建设,发展无所不在的网络和相关产业,并由此催生新一代信息科技革命。韩国信息通信产业部于2004年成立了“U-Korea”策略规划小组,提出了“U-Korea”战略,并在2006年确立了相关政策方针。2013年,韩国物联网产业规模已经达到50万亿韩元。

从2005年开始,我国学者开始借鉴国外智慧城市的建设与发展经验。近两年来,智慧城市获得了我国不少地方政府的认同。目前,北京、上海、深圳、无锡、武汉、南京等城市已纷纷启动智慧城市战略。这些城市的相关规划、项目和活动渐次推出,进入了我国智慧城市的第一梯队。上海、天津、无锡、深圳、沈阳、武汉、成都等地已分别建立了RFID(Radio Frequency Identification,无线射频识别技术)产业园区,期望能率先发展物联网产业。

智慧城市的未来趋势包括绿色宜居、平安有序、幸福和谐、集约高效和科学决策5个方面。所谓绿色宜居,有两层含义,一层是“绿色”,一层是“宜居”。绿色是手段,宜居是目的。这里的绿色又有两层含义,一是指狭义的绿化与绿化面积,另一个是广义的绿色,即环保、

节能、无害等。城市平安是对城市运行的最低要求,一座城市文明程度高低最重要的指标之一就是这座城市的秩序如何。所谓平安有序,平安是指通过智慧城市的智能化信息技术,如安防网络等,来保证整个城市居民的生活场所安全。所谓幸福和谐,人类社会发展到今天,从根本上来讲,仍是要解决人与自然的可持续发展、人与人的协调发展以及人对自我的认知与幸福感提升,而智慧城市就是用现代化科技手段去更好地协调处理好这三个关系。智慧城市的目标是幸福指数最大化,提高市民的生活幸福感。所谓集约高效,在城市资源越来越紧缺的今天,为了节约资源、保护环境,许多城市开始对其中心区建设进行深度思考和改造实践,希望寻找到更集约、更高效的城市发展模式,城市集约化是未来的一个发展趋势和方向。智慧城市为科学决策提供了支撑,智慧城市在其运行中将信息通信技术应用到城市运营和发展的方方面面,通过计算机化的系统结构有效完成感知、处理、决策等复杂行为,从而使得城市环境更加宜居、城市运营更加智能、城市管理更加高效。

1.1 智慧城市的起源、概念和特征

1.1.1 智慧城市的起源

20 世纪以来,世界人口急速城市化,城市人口比例由 1900 年的 13%(2.2 亿)增加至 1950 年的 29%(7.32 亿),2007 年已有 33 亿人生活在城市,超过了全球人口总数的 50%。有数据显示,未来亚洲和非洲的城市人口增长率高达 93%;2030 年,世界城市人口比例会增至 60%(50 亿);预计到 2050 年这一数字将升至 70%,全世界三分之二的人口(超过 60 亿)将居于城镇(摘自中国对外友好协会会长李小林 2011 年 12 月在“2011 沃特金融峰会”上的讲话)。由此可见,城市化是全世界未来发展的一种必然现象,城市在世界上的地位越来越重要。随着城市的发展,全球所面临的城市问题越来越多,也越来越突出。交通堵塞、环境污染、食品安全、资源短缺、地域发展不均衡、就业压力、通货膨胀等,无不影响着人们生活水平的提高,这就是近年来所说的“城市病”。要解决城市病问题,需要用新模式、新理念和新技术来支撑城市的发展,于是一种新型的城市建设模式——智慧城市,在各个国家不断的摸索中诞生了。

智慧城市一词最早于 1984 年由美国拉斯维加斯一家以“智慧城市”命名的产业技术协会组织提出。1992 年,美国 13 个州的城市兴起智慧增长运动,该运动主要由环保主义者发起,强调土地集约使用,经济重要性让位于环境和生活方式。2007 年,欧盟在《欧盟智慧城市报告》中提出了智慧城市的创新构想,而对这一构想的深入解读始于美国 IBM 公司总裁兼首席执行官彭明盛的一次演讲。2008 年 11 月 6 日,彭明盛在纽约市外交关系委员会发表了题为《智慧地球:下一代的领导议程》的演讲,提出了“智慧地球”的概念,即通过互联网

(Internet),把无处不在的、被植入城市物体的智能化传感器连接起来,最终实现人们对物理城市的全面感知,并调控城市各个方面资源,达到可持续健康发展的目的,为人类创造更美好的城市生活。具体措施是把感应器嵌入并装备到电网、铁路、桥梁、隧道、公路、建筑、供水系统、大坝、油气管道等各种物体中,并且彼此连接,形成所谓“物联网(the Internet of Things)”,然后将物联网与现有的互联网整合起来,实现人类社会与城市物理系统的整合和互动。智慧城市是 IBM 提出的智慧地球策略中的一个重要方面,目的是以更智慧的方法,通过新一代信息技术改变人们交互的方式,提高实时信息处理能力及感应与响应速度,增强业务弹性和连续性,促进社会各项事业的全面和谐发展(图 1-1)。

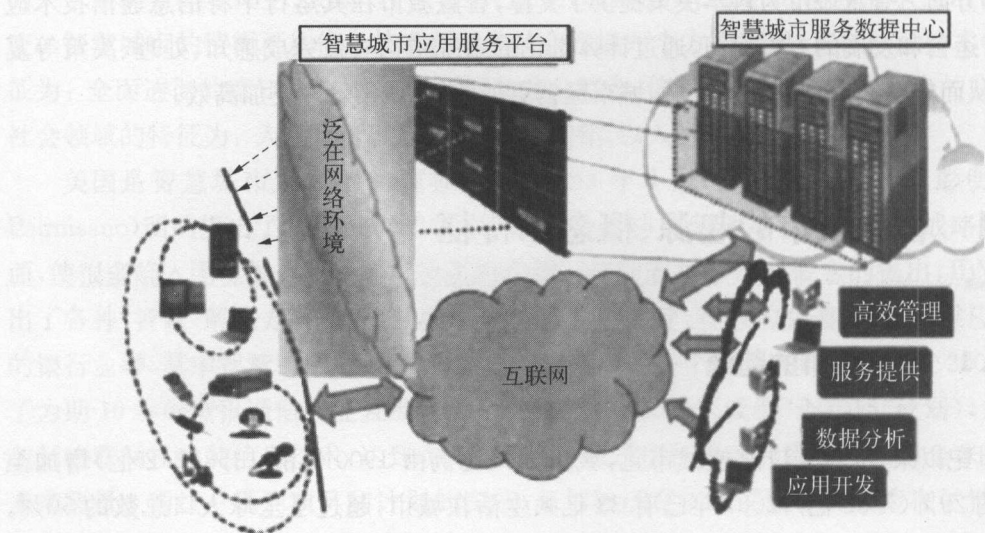


图 1-1 智慧城市信息系统

2009 年,奥巴马就任美国总统后与美国工商业领袖举行了一次圆桌会议。会上,彭明盛再次提出了智慧地球的理念,建议新政府投资建设下一代智慧型基础设施。奥巴马政府对此给予了积极回应,并将其上升为国家战略。

智慧城市是新一代信息技术支撑、知识社会创新环境下的城市形态,是信息化、工业化和城镇化的深度融合,它不仅会改变居民的生活方式,也会改变城市的生产方式,保障城市可持续发展。建设和发展智慧城市,是经济增长的“倍增器”和经济发展方式的“转换器”,是提高城镇化质量,缓解当前日益严重的“大城市病”的重要举措。

伴随着网络的崛起、移动技术的融合发展以及创新的民主化进程,智慧城市是继“数字城市(Digital City,美国副总统戈尔于 1998 年 1 月 21 日提出的概念)”之后城市信息化发展的高级形态。建立智慧城市的必要性主要表现在以下 3 个方面^[1]:

(1) 智慧城市是城市进入 21 世纪网络化信息时代,提升核心竞争力的战略需求。进入信息化网络时代后,人们面临的数据很多、处理很慢、自动化与智能化程度很低、知识很少的问题日益突出,这些问题的解决必然推动新一代信息技术的发展及其产业化步伐的加

快。智慧城市深刻反映了信息技术全面进入城市实体经济所带来的巨大变革趋势,是城市在 21 世纪提升核心竞争力的关键所在。

(2) 智慧城市是数字城市进一步深化发展的战略需求。数字城市的出现,有其社会背景和技术背景。但相对于国家信息化战略而言,数字城市只是起点不是终点,还需要深化发展。进一步解决数字城市关于网上城市“数字空间”与现实城市“物理空间”相分离的问题,只能走智慧化之路,即通过传感器网络和物联网把网上城市数字空间与现实城市物理空间“缝合”在一起,使数字城市进一步深化发展。因此,数字城市必然走向智慧城市。

(3) 智慧城市是实现智慧地球的战略需求。智慧地球的核心思想是无处不在的智能对象被无处不达的网络与人连接在一起,被无所不能的超级计算机调度和控制。这是一个全球性的战略目标,而它的具体实现依赖于智慧城市的建设。如同数字城市是数字地球的重要组成部分一样,智慧城市是智慧地球的战略需求。

智慧城市的发展对中国具有重要意义。诺贝尔经济学奖得主斯蒂格利茨说过这样的一句话:“中国的城市化和新技术革命是影响 21 世纪人类进程的两大关键性因素。”城市化和信息化的发展交融在一起,一定会形成城市发展的新模式。我国新一代的信息技术有互联网、云计算、无线网络等,当新技术的发展和城市化的进程结合到一起的时候,就出现了我国城市发展的新思路:一是实现人与人、人与物、物与物之间的连接,用无障碍的通信来提供完善的基础设施,这是我们发展最基本的思路;二是助力城市以低碳和环保方式发展,提供和谐宜居的生活环境,创造绿色城市生活;三是提高城市的运转效率,促进城市的可持续发展,打造高效城市管理和运行;四是促进城市产业结构的升级和知识型人才的聚集,加速城市化的进程。在这个新思路的指引下会形成一个新的模式,这就是智慧城市。推进我国智慧城市建设有利于推进内涵型城镇化发展;有利于培育和发展战略性新兴产业,创造新的经济增长点;有利于促进传统产业改造升级、社会节能减排,推动经济发展方式转型;有利于我国抢抓新一轮产业革命机遇,抢占未来国际竞争制高点^[2]。

1.1.2 智慧城市的概念

关于智慧城市,目前没有公认的统一定义。简单来说,智慧城市就是让城市更聪明,本质上是让作为城市主体的人更聪明。它通过互联网将无处不在的被植入城市各物体的智能传感器连接起来,实现对现实城市的全面感知,利用云计算等智能处理技术对海量感知信息进行处理和分析,实现网上城市数字空间与物联网的融合,并发出指令,对包括政务、民生、环境、公共安全等在内的各种需求作出智能化响应和智能化决策支持^[1]。它是信息化向更高阶段发展的表现,实现的关键是借助新一代的云计算、物联网、决策分析优化等信息技术,通过感知化、互联化、智能化的方式,将城市中的物理基础设施、信息基础设施、社会基础设施和商业基础设施连接起来,成为新一代的智慧化基础设施,使城市中各领域、各

子系统之间的关系显化,犹如给城市装上网络神经系统,使之成为可以指挥决策、实时反应、协调运作的“系统之系统”^[3]。

智慧的城市意味着更透彻的感知、更广泛的互联互通、更深入的智能化。也意味着在城市不同部门和系统之间实现信息共享和协同,更合理地利用资源、作出最优化的决策、及时预测和应对潜在风险、突发事件和各类灾害;目的是协调各职能部门,整合优化现有资源,提供更好的服务,为城市各类主体营造一个良好的工作、生活和休闲环境^[4]。

关于智慧城市的概念,不同的学者和机构从不同的角度给出了不同解释,现列举如下:

(1) IBM在《智慧的城市在中国》白皮书中,将智慧城市定义为这样一个城市:“能够充分运用信息和通信技术手段感测、分析、整合城市运行核心系统的各项关键信息,从而对于包括民生、环保、公共安全、城市服务、工商业活动在内的各种需求作出智能的响应,为人类创造更美好的城市生活。”

(2) 两院院士李德仁认为:智慧城市是在城市全面数字化基础之上建立的可视、可量测、可感知、可分析、可控制的智能化城市管理与运营机制,包括城市的网络、传感器、计算资源等基础设施,以及在此基础上通过对实时信息和数据的分析而建立的城市信息管理与综合决策支撑等平台。

(3) 工程院院士许庆瑞^[5]认为:智慧城市早期的概念大多是技术导向型的,属于智能城市范畴,注重的是城市建设的硬实力。后来进一步发展,关注的焦点不再局限于硬件设施的建设,更注重城市的人文和教育环境建设,即重视城市建设的软实力。

(4) 秦洪花(青岛市科技信息研究所)等认为,智慧城市即利用新一代信息技术,以整合、系统的方式管理城市的运行,让城市中各个功能彼此协调运作,为城市中的企业提供优质的发展空间,为市民提供更高的生活品质。智慧城市需要更加智能的城市规划和管理、资源分配更加合理和充分、城市有可持续发展的能力、城市的环境保护到位、能够提供更多的就业机会、对突发事件具备应急反应能力等。

(5) 徐国强(上海市城市规划设计研究院)等认为,智慧城市的本质含义是以城市可持续发展为目标,将先进信息技术与先进城市经营服务理念有机融合,通过对城市的地理、资源、环境、经济、社会等系统进行数字网络化管理,包括对城市基础设施、基础环境、生产生活相关产业和设施的多方位数字化、信息化的实时处理与利用,构建以政府、企业、市民三大主体的交互、共享平台,为城市治理与运营、城市公共管理与服务提供简捷、灵活的决策工具与服务模式,充分发挥空间信息承载应用技术手段的巨大潜力,以精细化管理理念大幅提升现代化城市管理与服务水平,推进城乡一体化协调发展^[6]。

(6) 张永民(陕西省咸阳市政府办公室)认为,智慧城市把城市里分散的、各自为政的信息化系统、物联网系统整合起来,提升为一个具有较好协同能力和调控能力的有机整体,这是以前所没有的,是传统意义上的城市信息化和数字城市的升华和飞跃,并被赋予了新的内涵。从智慧城市的基本内涵来说,智慧城市是新一轮信息技术变革和知识经济进一步发展的产物,是以互联网、物联网、电信网、广电网、无线宽带网等网络的多样化组合为基础,

更加广泛深入地推进基础性与应用型信息系统开发建设和各类信息资源开发利用,形成的技术集成、综合应用、高端发展的网络化、信息化、智能化和现代化城市,是以智慧技术、智慧产业、智慧人文、智慧服务、智慧管理、智慧生活等为重要内容的城市发展的新模式。简单地说,智慧城市是充分利用信息化相关技术,通过监测、分析、整合以及智慧响应的方式,综合各职能部门,整合优化现有资源,提供更好的服务、绿色的环境、和谐的社会,保证城市可持续发展,为企业及大众建立一个优良的工作、生活和休闲的环境。形象地说,智慧城市=智慧+互联+协同。

综上所述,智慧城市是一个复杂的社会大系统,通过自动传感、互联网、物联网、云计算等信息技术和自动化技术等手段,使构成城市的基础设施、环境等要素与人相互之间实现互联、协同和智能,作为一个系统整体对民生、发展、环保、公共安全、城市服务、经济活动等需求作出智慧的响应。

智慧城市包括以下内涵:

(1) 智慧城市的手段。智慧技术是智慧城市的手段,是对现有互联网技术、传感器技术、智能信息处理等信息技术的高度集成,是实体基础设施与信息基础设施的有效结合,是信息技术的一种大规模普适应用,将引发新一轮大规模的技术创新。

(2) 智慧城市的基础。有关信息的感知、获取、传递、控制、处理和利用的设施构成了智慧城市的基础。借助于信息技术,城市系统中的物理基础设施、信息基础设施、社会基础设施和商业基础设施连接起来,成为智慧化的基础设施^[4]。

(3) 智慧城市的核心。智慧城市是以互联网、物联网、电信网、广电网、无线宽带网等网络组合为基础的,其中,物联网体系的建设是智慧城市的核心。智慧城市的理念实质上是借助于物联网和互联网整合,以一种更智慧的方法感测、分析、整合城市运行核心系统的各项关键信息,改变政府、企业和人们相互交往的方式,对各种需求作出快速和智能化的响应,从而提高城市系统的智慧水平^[4]。

1.1.3 智慧城市的特征

从智慧城市的概念和内涵出发,建立和完善智慧城市的愿景是:整个城市具有较为完善的行为意识和调控能力,具有多平台协同能力;具有智能感知、情境感知与认知能力;具有成熟的信息—知识—智能转换机制和一定的决策能力;具有一定的自我学习、自我成长和自我创新能力(试点和建设开始后还会有新的认识)等。最近,上海浦东智慧城市发展研究院在其发布的“智慧城市指标体系 1.0 版”中指出,智慧城市是指综合利用各类信息技术和产品,以“数字化、智能化、网络化、互动化、协同化、融合化”为主要特征,通过对城市内人与物及其行为的全面感知和互联互通,大幅优化并提升城市运行的效率和效益,实现生活更加便捷、环境更加友好、资源更加节约的可持续发展的城市。智慧城市的特征可以从信息技术领域和社会学领域分别进行归纳。

1) 信息技术领域

目前,世界各国在智慧城市方面的探索与实践,都注重从市民需求出发,通过新一代信息技术的应用,以各种基础网络为支撑建设感知设施,通过信息的融合分析提供智能服务。在信息技术领域,智慧城市显现出4大基础特征:

(1) 全面透彻的感知。

① 广泛嵌入传感设备:要实现全面、透彻的感知是一项非常艰巨的任务,传感技术和设备的发展是关键。传感设备在智慧城市中的广泛嵌入形成了智慧城市的“感觉器官”,通过传感技术,实现对城市管理各方面的监测和全面感知。智慧城市利用各类感知设备和智能化系统,智能识别、立体感知城市环境、状态、位置等信息的全方位变化,对感知数据进行融合、分析和处理,并与业务流程智能化集成,继而主动作出响应,促进城市各个关键系统和谐高效的运行^[7]。

② 构建基础数据库:通过分布在城市重要基础设施、城市公共环境中的传感系统以及自动监测、监控设施的联网,实现城市海量信息与数据的实时收集与存储,构建个人信息、法人信息、地理信息、统计信息四大城市基础数据库,以及城市重大基础设施智慧监测信息、治安与道路实时监测信息等城市应用数据库。并在此基础上通过使用传感器、先进的移动终端、高速分析工具等,实时分析城市中的所有信息,以便政府及相关机构及时作出决策并采取适当措施,从而实现高智能化和全面感知^[3]。

(2) 物联与互联网的完全整合。

① 宽带泛在互联:各类宽带有线、无线网络技术的发展为城市中物与物、人与物、人与人的全面互联、互通、互动,为城市各类信息随时、随地、按需、随意应用提供了基础条件。宽带泛在网络作为智慧城市的“神经网络”,极大地增强了智慧城市作为自适应系统的信息获取、实时反馈、随时随地提供智能服务的能力^[8]。城市各要素通过高带宽的固定网络、无线网络、移动通信网络等,实时在线地连接起来,使人们能从全局的角度分析并实时解决问题,从而彻底改变城市管理与运作的方式。

② 全面物联:物联网是通过信息传感设备,按约定的协议将任何物品与互联网进行连接以及信息交换和通信,从而实现智能化识别、定位、跟踪、监控和管理的一种网络。通俗而言,即是把物品与物品通过网络连接起来,使人可以通过网络识别方法控制各类物品。物联网作为建设智慧城市的关键技术,如当年互联网一样具有推动新一轮全球科技和经济浪潮的潜力。世界各主要经济体纷纷将发展物联网作为应对金融危机、扩大就业、抢占未来制高点的重要战略。

③ 两者充分整合:物联网与互联网系统完全连接和融合,将数据整合为城市核心系统的运行全图,借助于发达的智慧基础设施,提升信息处理和信息资源整合能力。

(3) 信息可靠传递。

① 广泛联结:移动互联网、电信网、互联网、物联网彼此融合形成泛在化的网络承载系统,安全可靠地将各种采集信息和控制信息进行实时准确的可靠传递。基于广泛联结的可

靠传递是智慧城市信息来源的基础,广泛联结如同智慧城市的“经络”,而可靠传递如同来自外部的准确“刺激”,实现智慧城市对外界信息的准确通信^[9]。

② 交换共享:通过管理体制的创新保障,构建身份认证、目录交换、结算清算、信用评估等技术平台的体系性建设,确立信息系统之间的层次性,促进分布在城市不同角落海量数据的流转、交换、共享、比对,为信息应用提供良好的协同工作环境^[3]。

(4) 智能融合的应用。

① 更加集中、更有深度计算的智能处理能力:现代城市及其管理是开放的复杂巨系统,新一代全面感知技术的应用增加了城市的海量数据。对海量数据进行智能化的处理是实现智慧城市的关键和标志,是智慧城市区别于数字城市的关键点。基于云计算、数据挖掘、智能模糊识别等各种智能计算技术,对海量数据进行快速、集中、准确的分析和处理,并做出智能化的控制与处理,并引入综合集成法,通过人的“智慧”参与,提升决策支持和应急指挥的能力,将构成智慧城市的“大脑”^[8]。

② 技术融合与发展:进一步推动“云”与“端”的结合,推动从个人通信、个人计算到个人制造的发展,推动实现智能融合、随时、随地、按需、随意的应用,进一步彰显个人的参与和用户的力量^[8]。

③ 关联应用:在互联互通网络、数据交换与共享基础上,以政府、市民、企业的互动为核心,构建公共管理与服务平台,可以提供整合式的协同服务,从而为城市管理与运营提供更智能高效、响应更灵活及时的决策支持系统、管理服务手段和创新应用模式,使城市达到最佳运行状态。

此外,为了更好地诠释智慧城市“智”的特征,中国工程院院士王家耀^[1]将“智”概括为以下6个方面:

(1) 透彻感知。无处不在的智能传感器,对城市物理空间实现全面、综合的感知,对城市运行的核心系统实时感测,智能化获取现实城市物理空间的各种信息。

(2) 全面互联。通过无处不在的互联网将无处不在的智能传感器网络(包括卫星传感器网和物联网)连接起来,实现感知数据的智能传输、存储和管理。

(3) 深度整合。通过城市“三网”融合、互联网与物联网、卫星传感器网的融合和基于云计算平台的多源异构数据(多参考系、多语义、多尺度、多时相等)的整合(集成、融合与同化),构建智慧城市的信息基础设施。

(4) 协同运作。基于城市智慧信息基础设施,智慧城市的各要素(信息获取、处理与服务)、单元和系统及其参与者和谐高效地运行,达到城市运行的最佳状态。

(5) 智能服务。在城市智慧信息基础设施基础上,通过构建一种新的服务模式或一种新的能提供服务的体系结构,对海量感知数据进行处理、数据挖掘与知识发现,为人们提供各种不同层次、不同要求的低成本、高效率的智能化服务。

(6) 激励创新。鼓励政府、企业和个人在城市智慧信息基础设施上进行科技和业务的创新应用,寻求新的经济增长点,为城市经济、社会发展和文明进步提供源源不断的动力。

2) 社会学领域

很多学者认为,硬件设施建设是智慧城市的基础,但要想保证其可持续发展,离不开人文、教育和社会资本在其中的作用。国际上智慧城市的先行者都特别强调以市民为中心,强调用户参与、社会参与的开放创新空间构建。例如,欧盟在对中小城市开展的智慧城市评价中,梳理了智慧经济、智慧公众、智慧管理、智慧移动性、智慧环境、智慧生活 6 大维度、31 个方面、74 项指标,体现了智慧城市建设中以人为本、强化服务、强化价值创造的创新理念。归纳起来,在社会学领域,智慧城市显现出 3 大基础特征:

(1) 人性化管理与服务。智慧原本是对人的灵性的描述,现在移植到城市建设之中,其目的是要实现城市的智能化、自动化、智慧化、人性化等,即城市像人一样也有灵性和智慧。当城市的运行建立在全面感知、可靠传递以及智能处理的基础之上时,城市也如同人一般具有了灵性和智慧,这就是智慧城市^[9]。

(2) 协同运作。基于智慧的基础设施、城市里的各个关键系统和参与者同时进行和谐高效的协作,有利于提高跨部门、多层级、异地的合作能力,实现城市运行的最佳状态。

(3) 以人为本的可持续创新。面向知识社会的可持续创新重塑了现代科技以人为本的内涵,也重新定义了创新中用户的角色、应用的价值、协同的内涵和大众的力量。智慧城市的建设尤其注重以人为本、市民参与、社会协同的开放创新空间的塑造以及公共价值与独特价值的创造。它通过维基、微博、Fab Lab、Living Lab 等工具和方法强化用户的参与,汇聚公众智慧,不断推动用户创新、开放创新、大众创新、协同创新;鼓励政府、企业和个人在智慧基础设施之上进行科技和业务的创新应用,为城市提供源源不断的发展动力,以人为本地实现经济、社会、环境的可持续健康发展。

1.2 智慧城市的发展现状

1.2.1 智慧城市的发展

智慧城市既是社会需求的产物,也是技术推动的结果,从 19 世纪 90 年代至今,各国智慧城市发展迅速。全球新一代信息技术的快速发展,为实现从数字城市到智慧城市的转变提供了可能。主要表现在以下 4 个方面:

(1) 数字城市建设为智慧城市的实现奠定了坚实基础。建设智慧城市,数字城市建设是先行和基础,没有数字城市就没有智慧城市,因为智慧城市是通过物联网把数字城市与现实城市融合在一起的产物。相对于数字城市的数据采集、处理与分析、应用与服务而言,智慧城市更加智慧化、实时化^[1]。

(2) 物联网为智慧城市战略提供了新一代信息基础设施。互联网实现了计算机的联

网,万维网(Web)实现了网页的联网,而物联网则是集传感器、无线射频识别技术、通信网与互联网技术、智能运算技术等感知技术于一体,实现以全面感知、可靠传输、智能处理为特征的,连接物理世界的网络。所以,物联网的出现和发展,为智慧城市的实施提供了强大的技术和物质支撑。2009年11月,在全球物联网会议上,欧盟专家介绍了《欧盟物联网行动计划》,意在引领世界物联网产业发展。在欧盟,较为活跃的各大运营商和设备制造商推动了M2M(Machine to Machine,机器与机器)的技术和服务的发展。为了加强政府对物联网的管理,消除物联网发展的障碍,欧盟制定了一系列物联网管理规则,并建立起一个有效的分布式管理架构,使全球管理机构可以公开、公平、尽责地履行管理职能。为加深各相关方对物联网机遇、挑战的理解,欧盟同时还致力于增强机构间的协调,共同推动物联网发展,欧盟执行委员会定期向欧洲议会、欧盟理事会、欧洲经济与社会委员会、欧洲地区委员会等相关机构通报物联网发展情况^[10]。

(3) 云计算为智慧城市提供了新的应用服务模式。云计算并不是单纯的技术方法,而是方法论,是思维方式和工作方法的改变,是一种新的服务方式。它以应用为目的,通过互联网将大量的硬件、软件和数据等资源按照一定的结构体系连接起来,并随应用需求的变化不断调整结构体系,建立内耗最小、功效最大的虚拟资源服务中心,组成一个庞大的资源池,并通过网络传输给用户^[1]。对智慧城市而言,云计算解决了以下3个问题:一是基于云计算公共服务平台的多源结构数据集成、融合与同化,因为“云”里的数据是连通的,不能存在数据“孤岛”;二是基于云计算公共服务平台支撑的动态增量数据的存储与管理,因为智慧城市面临的物联网数据是动态增量的;三是基于云计算公共服务平台的软件服务,根据应用需求的变化,动态调整、修改软件,并实现系统的集成。

(4) 目前出现的初级智慧城市应用案例具有一定的技术积累。从数字城市到智慧城市是一个发展过程,两者之间既有联系也有区别。一些国家和地区已出现的某些智慧城市应用案例,如“智慧交通”领域的“自动收费”、“智慧医疗”领域的“个人保健”、“医疗监护”等,在关键技术方面已取得了某些突破,这些都为实现从数字城市到智慧城市积累了经验和

技术。

近年来,智慧社区论坛(Intelligent Community Forum)的成立和评比活动对智慧城市理念的普及和推广产生了积极的影响。它每年都举办全球的智慧城市评比。评比分3个阶段:第一阶段是从全球众多城市中遴选出 Smart 21;第二阶段,再从中选出 Top 7;最后评选出年度智慧城市(Intelligent Community of the Year)。第一届世界智慧城市评比自1999年开始,至今已连续举办了14届(2003年除外)。通过国际智慧城市评比,世界很多国家和地区认识到了智慧城市的前瞻性和先进性,相继提出了建设智慧城市的战略举措。据统计,目前全球已有1200多个智慧城市项目。

在我国,目前智慧城市发展的产业基础和技术基础基本形成,政策环境不断改善,初步具备了深入发展的基本条件,主要体现在以下几个方面:

(1) 城市信息化升级的基础条件基本具备。近年来,我国信息化进程正步入快车道,数

字城市建设试点和推广城市已达130个,近60个城市基本完成建设,国家信息化水平指数从2002年的0.6487上升到2009年的1.5458^[8]。据信息化蓝皮书《中国信息化形势分析与预测(2011)》数据显示,2006~2008年,我国信息化发展指数年均增长13.30%,居世界第5位,高出世界平均增速1倍,在信息化可接入性和可使用性方面更是全球进步最快的国家。我国各城市高度重视信息基础建设,无线通信网络和宽带覆盖率等信息化指标显著提升,政务、商业、交通、医疗、教育等领域的信息化水平不断提高^[1]。

(2) 智慧技术链和产业链初步形成。从智慧技术与产业发展上看,我国智慧技术产业化和信息产业融合进程不断加快,产业规模效应初步呈现。我国是世界电子信息产业的重要生产基地,产业规模位居世界第二。以物联网为例,我国物联网技术研发水平已经处于世界前列,在无线智能传感器网络通信技术、微型传感器、传感器端机、移动基站等方面都已取得重大进展,已成为国际标准制定的主导国之一。我国目前已落实了250多个物联网试点项目,建立了江苏无锡国家传感网示范区,北京、上海等5个国家云计算服务创新示范城市。我国基本建成了从材料、技术、器件、系统到网络的物联网产业链,并呈现出带动电信运营商、高校、科研机构、传感器企业、系统集成商、应用软件开发商等环节聚合联动之势,产业规模效应不断外溢。与此同时,在相关核心技术研发与产业化方面成绩显著,信息产业由硬件和制造主导向软件和服务主导转变的速度加快,系统集成和产业融合已成为产业创新发展的主要模式,信息产业内部交叉融合、重组融合以及与其他产业的渗透融合不仅拓展了信息技术的应用领域,而且派生出许多新兴业态,进一步壮大了信息产业规模^[1]。

(3) 从中央到地方的政策支持体系不断完善。国家高度重视信息技术和经济社会信息化发展,出台了《国家中长期科学和技术发展规划纲要(2006—2020)》、《国家信息化领导小组关于我国电子政务建设指导意见》、《国务院关于加快培育和发展战略性新兴产业的决定》、《进一步鼓励软件产业和集成电路产业化发展的若干政策》等一系列促进城市信息化发展的政策文件,对我国信息化的发展战略、目标和任务作出了明确规定。在“十二五”时期,智慧城市与集成电路、智慧工业、地理信息、软件信息服务等共同成为国家“十二五”规划以及“十二五”科技计划的重点,成为信息化带动工业化战略的重要着力点。在地方,许多城市将智慧城市建设纳入地方“十二五”规划纲要,先行地区还出台了一系列专门的政策规划,逐步建立起保障智慧城市稳定健康推进的政策体系。

2009年,国家总理温家宝在北京科技界大会上作了题为《让科技引领中国可持续发展》的报告,报告中诠释了物联网、智慧地球等智慧城市的概念,标志着智慧城市的研究已引起国家层面的重视。当年2月24日,IBM在中国提出“智慧地球赢在中国”。并建议优先建设智慧的电力、智慧的医疗、智慧的城市、智慧的交通、智慧的供应链(物流)、智慧的银行等六大行业。当年8月,IBM发布了题为《智慧的城市在中国》的报告,报告认为:有效利用信息技术提升城市管理水平,推动中国社会的城市化进程,已经成为城市管理者的当务之急。

2010年上海世博会期间,在上海举办了“2010智慧城市全球峰会”。IBM借助该时机,

积极倡导和推广智慧城市的发展理念和建设方案。6月2日,彭明盛发表“从城市开始构建智慧的地球”的主题演讲,认为:世界将继续变得越来越平、越来越小、越来越普遍联系在一起。世界也在变得越来越智慧,所有这些变化都会集中在城市呈现出来。全球峰会在IT界、经济界和学术界引起了巨大反响,引起社会各界的广泛关注和高度认同。11月2日,由科技部等单位举办的“2010中国智慧城市论坛”在武汉召开。论坛的主题是“发展更科学,管理更高效,社会更和谐,生活更美好”。12月12日,以“加速物联网发展与应用,共享智慧城市建设经验”为主题的2010中国物联网与智慧城市建设高峰论坛在北京召开。当年,中国智慧工程研究会制定完成了《中国智慧城市(镇)建设行动纲要(建议案)》,提出未来5年发展100个智慧城市(镇)、200个智慧城区示范区的建设构想。此外,智慧城市指标体系进入了研究者视野,中国智慧工程研究会发布了“中国智慧城市(镇)科学评价指标体系”、上海浦东新区发布了“智慧城市指标体系1.0”。

2011年5月18日,第二届APEC智慧城市智能产业高端会议在河北省廊坊市举行,智慧社区论坛的评审团主席约翰·荣格应邀参加。

1.2.2 国外智慧城市现状

1) 美国

作为世界第一经济强国,美国2008年尽管遭受了惨重的世界金融危机冲击,但丝毫不影响其在新市场方面的计划。2009年1月28日,奥巴马就任美国总统后,与美国工商业领袖举行了一次“圆桌会议”。作为仅有的两名代表之一,IBM首席执行官彭明盛明确提出了智慧地球概念,希望通过加大对宽带网络等新兴技术的投入,振兴经济并确立美国的未来竞争优势。奥巴马积极回应IBM的智慧地球概念,将智慧地球作为保持和重夺国家竞争优势的根本所在:“经济刺激资金将会投入到宽带网络等新兴技术中去,毫无疑问,这就是美国在21世纪保持和夺回竞争优势的方式。”美国政府将智慧地球上升到国家战略及政策层面,使很多陷入困境的企业看到了全新的希望。IBM智慧地球战略的主要内容是:把新一代IT技术充分运用在各行各业之中,即把感应器嵌入和装备到全球每个角落的医院、电网、铁路、桥梁、隧道、公路、建筑、供水系统、大坝、油气管道等各种物体中,通过互联形成物联网,而后通过超级计算机和云计算将物联网整合起来,使人类能以更加精细和动态的方式管理生产和生活,从而达到全球智慧状态,最终形成“互联网+物联网=智慧的地球”的状态^[11]。

伴随着智慧地球概念的提出,IBM相继推出了各种智慧解决方案,包括智慧的电力、智慧的医疗、智慧的交通、智慧的供应链、智慧的银行业等,其中智慧城市是IBM智慧地球策略的一个重要方面。构建智慧的地球从城市开始,智慧城市是智慧地球的缩影。2009年9月,美国中西部艾奥瓦州的迪比克市与IBM共同宣布,将建设美国第一个智慧城市——一个由高科技充分武装的60000人社区。通过采用一系列IBM新技术武装的迪比克市将完

全数字化,并将城市的所有资源都连接起来(水、电、油、气、交通、公共服务等),有关部门可以侦测、分析和整合各种数据,并智能化地做出响应,服务于市民的需求,以期打造更加节能,智能化的城市^[11]。

2010年3月,美国联邦通信委员会(FCC)正式对外公布了未来10年美国的高速宽带发展计划,将目前的宽带网速度提高25倍,到2020年以前,让1亿户美国家庭互联网传输的平均速度从现在的每秒4 Mb提高到每秒100 Mb。而此前的20世纪90年代,克林顿政府也曾耗资2 000亿~4 000亿美元,用20年时间建成美国国家信息高速公路基础设施,创造了巨大的经济和社会效益,使美国成为全球信息产业强国。近年来,美国又在7 870亿美元的《经济复苏和再投资法》中提出,从能源、科技、医疗、教育等方面着手,通过政府投资、减税等措施来改善经济、增加就业机会,带动美国长期发展,其中鼓励物联网技术发展的政策主要体现在推动其在能源、网络宽带与医疗三大领域的应用。无论从基础设施、技术水平,还是产业链发展程度看,美国在新一轮技术创新浪潮中将走在世界各国的前列,趋于完善的互联网络为其物联网的发展创造了良好的先机。

2) 新加坡

2006年,新加坡推出了为期10年的资讯通信产业发展蓝图——“智慧国2015”(或称“iN2015”计划),力图通过包括物联网在内的信息技术,达到成为一个由资讯通信所驱动的智慧国家与全球型都市的未来愿景。“智慧国2015”计划中,提出了一个很有意义的概念,即要通过智慧化过程,在一些公共服务领域实现从供给方主导向供给方与需求方双向互动的转变,实现从非连续、碎片化的服务向连续性、一体化的服务转变。

通过一系列项目和计划的实施,新加坡已在物联网建设方面走在了世界前列。例如:在新加坡,平均每人拥有1.6辆机动车,城市交通压力极大,由于交通拥堵导致了效率降低和环境破坏,每年因此造成的损失占GDP的1.5%~4%。早在1998年,新加坡陆路交通管理局就开始着手建造电子道路收费系统(Electric Road Pricing System, EPS),通过对道路交通数据的收集和测算来界定拥堵路段,汽车在交通拥堵路段通行时要进行收费。这一做法当时在世界上还是首创。据该交通管理局报告称,相比交通高峰时期,实施后的道路通行量减少了25 000辆汽车,车流量提高了20%,取得了较好成效。但是,仅通过EPS来构建智能交通系统还是不够的,已经上马的智能交通系统(ITMS),将城市路网信息连接成网络,同时安装传感器、红外线设备,通过优化交通信号系统、电子扫描系统、城市快速路监控信息系统、接合式电子眼以及ERP系统等,提供历史交通数据和实时交通信息,对预先设定的时段(10 min、15 min、30 min、45 min和60 min)的交通流量进行预测,市民可以通过手机网络、车载GPS查询未来1 h内的交通情况,并选择合适的出行时间和路线。该交通管理局控制1 700个交通信号灯,对未来1 h内各个路段情况的平均预测准确率达到惊人的85%以上,10 min内的预测结果准确率更是高达90%^[11]。

经过多年发展,新加坡在利用信息通信技术促进经济增长与社会进步方面都处于世界领先地位,在信息通信项目、互联互通及电子政务方面取得的成绩更是引人注目。2008年,

新加坡仅在信息通信项目方面就投入约 16 亿新元(约 80 亿人民币),到目前为止,总投资额已经超过 40 亿新元,主要用于建立超高速、广覆盖、智能化、安全可靠的资讯通信基础设施以及发展从业人员的资讯通信技能。目前,信息通信每年为新加坡贡献约 6% 的国民生产总值。作为东南亚的重要航运枢纽,新加坡注重利用信息通信技术增强港口和各物流部门的服务能力,由政府主导,大力支持企业和机构使用 RFID 及 GPS 等多种技术增强管理和服务能力。此外,新加坡的电子政府公共服务架构(Public Service Infrastructure)已经可以提供超过 800 项政府服务,真正建成了高度整合的全天候电子政府服务窗口;其网络建设现有 130 万用户,其中 35% 的用户每周平均用网超过 3.6 h,迄今为止,网速达 1 kB/s 的新一代全国宽带网络已覆盖新加坡 35% 的房屋和建筑,并将于 2012 年实现 95% 的覆盖率;作为有线宽带的补充,“无线@新加坡”项目通过 7 500 多个热点,提供速度高达 1 Mb/s 的无线 Wi-Fi 上网服务,相当于每平方公里就有 10 个公共热点^[11]。

3) 瑞典

瑞典的智慧城市建设在交通系统上得到了最大的体现。智慧改变交通,IBM 和瑞典首都斯德哥尔摩的合作已经有了很好的示范。在斯德哥尔摩,四通八达的公共交通网络虽然能够为 70% 的上班族提供服务,但平均每天仍有 45 万辆汽车驶过城市中央商务区,交通严重拥堵时有发生。斯德哥尔摩市政府官员与瑞典公路管理局经过认真讨论,认为可以通过改造道路使用模式,将拥堵控制在一定范围内。要想改造道路使用模式,公路管理局需要准确动态地测量并且跟踪道路使用情况,据此收取使用费。2006 年初,瑞典当局宣布征收“道路堵塞税”,IBM 公司为公路局设计、构建并且运行了一套先进的智能收费系统,包含摄像头、传感器和中央服务器等。在 IBM 公司的帮助下,公路管理局在通往斯德哥尔摩市中心的道路上设置了 18 个路边控制站,通过使用 RFID 技术以及利用激光、照相机和先进的自由车流路边系统,自动识别进入市中心的车辆,自动向在周一至周五(节假日除外) 6:30~18:30 进出市中心的注册车辆收税。这一举措将斯德哥尔摩市的交通量降低了 20%,排放量减少 12%。瑞典当局通过收取“道路堵塞税”,减少了车流量,交通拥堵降低了 25%,交通排队所需的时间下降 50%,道路交通废气排放量减少了 8%~14%,二氧化碳等温室气体排放量下降了 40%。由于在环保方面做得出色,2010 年 2 月,斯德哥尔摩被欧盟委员会评为首个“欧洲绿色首都”^[12]。

4) 日本

日本在 2004 年推出了基于物联网的国家信息化战略,称作“U-Japan”。“U”指英文单词“Ubiquitous”,意指普遍存在的、无所不在的。该战略旨在推进日本信息通信技术建设,发展无所不在的网络和相关产业,并由此催生新一代信息科技革命。“U-Japan”由日本信息通信产业的主管机关总务省提出,目标是在 2010 年将日本建设成一个“任何时间、任何地点、任何人、任何物”都可以上网的环境,使日本成为一个充满朝气的国家,使所有日本人,包括儿童和残疾人,都能积极地参与日本社会的活动;通过无所不在的物联网,创建一个新的信息社会。为了实现“U-Japan”战略,日本进一步加强官、产、学、研的

有机联合,在具体的政策实施上,以“民、产、学”为主,政府的主要职责就是统筹和整合。通过实施“U-Japan”战略,日本希望开创前所未有的网络社会,并成为未来全世界信息社会发展的楷模和标准,在解决其高龄化等社会问题的同时,确保其在国际竞争中的领先地位^[13]。

2009年7月,日本政府IT战略本部又推出了至2015年的中长期信息技术发展战略“I-japan(智慧日本)战略2015”。该战略是日本继“U-Japan”之后提出的更新版本的国家信息化战略,其要点是大力发展电子政府和电子地方自治体,推动医疗、健康和教育的电子化。该战略旨在到2015年实现“以人为本,安心且充满活力的数字化社会”,让数字信息技术如同空气和水一般融入每一个角落,并由此改革整个经济社会,催生出新的活力,实现积极主动的创新。该战略的要点在于实现数字技术的易用性,突破阻碍其使用的各种壁垒,确保信息安全,最终通过数字化技术和信息技术向经济社会的渗透,打造全新的日本。战略由三个关键部分组成,一是设置“电子政务”、“医疗保健”和“教育人才”三大核心领域,二是激活产业和地域的活性,三是培育新产业及整顿数字化基础设施。为了体现以人为本,创造使国民安心和有活力的社会,“I-Japan”战略设置了一个核心内容——国民个人电子文件夹,目的是让国民管理自己的信息资料,通过互联网安全可靠地完成工资支付等各种手续。政府对其进行综合管理,使国民享受到一站式的电子政务服务。这一项目要求在2013年完成。日本政府希望,通过执行“I-Japan”战略,开拓支持日本中长期经济发展的新产业,大力发展以绿色信息技术为代表的环境技术和智能交通系统等重大项目。在上海世博会上,日本馆更是以“连接”为主题,用信息化最新科技让人们看到未来20~30年城市“智慧生活”的美好场景,例如展会上所亮相的“未来邮局”融合了互联网和物联网技术,在邮局中不仅能够寄送信件,还能实现人与商品的智慧交流^[11]。

5) 韩国

2004年,韩国信息通信产业部成立了“U-Korea”策略规划小组,提出了“U-Korea”战略,并在2006年确立了相关政策方针^[11]。“U-Korea”旨在建立“无所不在的社会”(Ubiquitous Society),即通过布建智能网络(如IPv6、BcN、USN)、推广最新的信息技术应用(如DMB、Telematics、RFID)等信息基础环境建设,让韩国民众可以随时随地享有科技智能服务。其最终目的,除运用IT科技为民众创造衣、食、住、行、体育、娱乐等各方面无所不在的便利生活服务之外,也希望通过扶植韩国IT产业,发展新兴应用技术,强化产业优势与国家竞争力。根据规划,“U-Korea”发展期为2006~2010年,成熟期为2011~2015年。

2009年10月,韩国通过了《物联网基础设施构建基本规划》,将物联网市场确定为新的增长动力。韩国通信委员会确定了到2012年的目标:通过构建世界最先进的物联网基础设施,打造未来广播通信融合领域超一流ICT强国。为了实现这一目标,韩国确定了构建物联网基础设施、发展物联网服务、研发物联网技术、营造物联网扩散环境等4大领域、12项详细课题。2013年,韩国物联网产业规模已经达到50万亿韩元^[13]。

此外,韩国仁川市2009年宣布与美国思科公司合作,以网络为基础,全方位改善城市管

理效率,努力打造一个绿化的、资讯化的、无缝连接的、便捷的生态型和智慧型城市。通过整合式的公共通信平台,以及无所不在的网络接入,消费者不仅可以方便地实现远程教育、远程医疗、远程办理税务事宜,还可以智慧化地控制房间的能耗等。

1.2.3 国内智慧城市现状

从2005年开始,我国学者开始借鉴国外智慧城市的建设与发展经验。2009年11月,温家宝总理在题为“让科技引领中国可持续发展”的讲话中,将物联网列入六大战略性新兴产业,国内随即掀起一股智慧城市、物联网热潮。在“十五”、“十一五”及“十二五”期间,中国国家“863”计划都有与智慧城市相关的重大项目开展,如“城市空间信息系统网格化集成与智能化服务技术”等。

近两年来,智慧城市获得了我国不少地方政府的认同。目前,除北京、上海以外,深圳、无锡、武汉、南京等国内城市也已纷纷启动智慧城市战略(表1-1),意在抢占先发优势。这些城市的相关规划、项目和活动渐次推出,进入了我国智慧城市的第一梯队。上海、天津、无锡、深圳、沈阳、武汉、成都等地已分别建立了RFID产业园区,期望能率先发展物联网产业。统计显示,2009年,中国RFID市场规模已达85.1亿元,同比增长29.3%,在全球居第三位,仅次于英国和美国。

表 1-1 我国主要的智慧城市建设情况

城 市	具 体 举 措
北 京	组建了中关村物联网产业联盟,建成 20 万亿次公共云计算平台
上 海	将在上海建设涵盖交通、医疗、安防、物流等各领域 10 个物联网应用示范工程
深 圳	首次提出“智慧深圳”理念,开始了由“知识”向“智慧”的进发,正从科技、人文、生态 3 个方面打造新时期的智慧城市
无 锡	在智能交通、健康工程、平安城市、环境保护等方面加快推进 TD 与传感网融合
武 汉	将用 10 年打造智慧城市,构建基于“中国云”的智慧城市基础设施及智能处理基础平台,建设 4 个智能示范应用工程,并力争突破 6 个关键技术
南 京	打造“智慧之都”、“绿色之都”、“枢纽之都”以及“博爱之都”
沈 阳	创新运用绿色科技和智慧技术,努力实现打造“生态沈阳”的目标
杭 州	提出建设“绿色智慧城市”,实现建设“天堂硅谷”和“生活品质之城”的城市发展战略目标
南 昌	提出把打造“数字南昌”作为智慧城市建设的突破重点
重 庆	提出要打造“健康重庆”

以下列举我国部分城市建设智慧城市的情况:

1) 无锡

无锡发展智慧城市的成果主要体现在物联网建设方面。由于2009年8月温家宝总理在无锡考察时提出了“感知中国”的理念,无锡在物联网发展方面走在了国内各大城市的前面。目前,中科院、清华大学、东南大学等20多家科研单位和高校均在无锡设立了研究机构,中国移动、中国联通、中国电信三大运营商也与无锡政府签约开展传感网络应用技术研究合作^[3]。为了更进一步推进物联网产业发展,无锡制定了详细的物联网产业规划,预计未来10年将投入330亿元资金,建设传感网络产业创新示范基地。

2) 南京

南京率先出台了智慧城市建设的专项规划——《南京市物联网产业发展规划》,明确了2010年南京市推出智能工业、智能环保、智能交通、智能灾害控制、智能农业、智能公共安全、智能医护、智能电网等物联网十大应用示范工程,这一系列举措昭示着南京开始加入智慧经济先机的争夺战^[13]。

3) 武汉

武汉2009年GDP为4560亿元,在内地副省级城市中名列第五,是华中地区最大的都市及中心城市,同时高校林立,科研院所众多,这为武汉发展物联网产业提供了有效的技术和人才支持^[13]。温家宝总理在无锡提出建设“感知中国”中心后,武汉10多家企业、科研单位联合成立了RFID创新技术联盟。2010年,武汉市政府与武汉大学开展合作,重点建设以“物联网研究院”等四大项目为代表的一批高起点、高水平、国际化的研究创新平台。

4) 深圳

当国内各城市高调进军物联网时,深圳迅速成立了RFID产业标准联盟,标志着深圳的RFID产业发展由技术应用向标准研制、由“企业单干”向“联盟协作”阶段转变^[13]。目前深圳已经集聚700多家相关企业,初步形成了包括RFID芯片设计与开发、电子标签制造、标签封装设备制造、读写器研发与制造、RFID应用软件开发、系统集成及咨询在内的产业生态链,其中不乏业界的龙头企业,如中兴通讯、华为、神州数码、远望谷、先施科技、惠田实业等。丰富的应用经验、庞大的物联网产业企业集群使得深圳成为了中国物联网发展的“黄埔军校”,而且深圳拥有雄厚的经济基础和产业基础,加之政府积极支持,将率先建设具有自身特色的物联网。

5) 重庆

重庆不但是直辖市,也是五大国家中心城市之一,未来将承担国内西部发展物联网产业的重任。重庆于1999年提出建设信息港,2000年提出建设“数字重庆”,但是还没有出台针对物联网产业发展的专项规划,其物联网产业发展的助推力更多是来自各大企业。早在2006年,中国移动全国M2M运营中心就落脚重庆,形成了电梯监控、车辆控制、企业安防三大成熟应用产品,并已推向市场商用,发展出7万多终端用户^[13]。

1.2.4 上海智慧城市现状

上海是拥有 2 000 万以上人口的特大型城市。在经历了一段时期超常规的快速发展后,上海在城市经济与产业发展、城市安全与管理、城市生活与公共资源配置、城市公共服务与社会保障、城市文化与社会进步等各方面面临很多问题,如资源紧缺、人口膨胀、饮水卫生、安全隐患、环境污染和交通拥挤等。2008 年,上海 GDP 增长仅 8.7%,终结了连续保持 16 年之久的两位数经济增长纪录,并低于同年全国 9% 的平均增速。2009 年,上海 GDP 增长 8.2%,继续低于全国平均水平的 8.7%。要解决上述问题,无论是历史上还是当今世界上都无现成的模式可以照搬。在增长乏力的现实下,上海急需寻求一种可持续的健康发展模式,找到新的城市增长动力。智慧城市建设涉及交通、通信、水、能源、医疗、教育等与城市功能相关的各个领域,能对城市管理、功能转型带来极大的促进作用,同时具有巨大的产业拉动作用,因此,城市智慧化是上海经济与社会发展的必由之路。

2010 年 11 月 9 日,上海市委通过了《关于制定上海市国民经济和社会发展的第十二个五年规划的建议》,其中把信息化提升到新的高度,提出“大力实施信息化领先发展和带动战略”,加快建设以数字化、网络化、智能化为主要特征的智慧城市,进一步促进整个城市“创新驱动、转型发展”。上海智慧城市的建设目标总体概括为:重点推进信息网络综合化、宽带化、物联化、智能化,充分发挥城市智慧型产业优势,加快商务、文化教育、医药卫生、城市建设管理、城市交通、环境监控、公共服务、居家生活等领域的智慧型建设,全面提高资源利用效率、城市管理水平和市民生活质量,将城市建设成为基础设施先进、信息网络通畅、科技应用普及、生产生活便捷、城市管理高效、公共服务完备、生态环境优美、惠及全体市民的智慧城市。

多年来,上海的城市信息化水平始终处于国内领先地位,为建设智慧城市创造了良好的基础条件。20 世纪 90 年代初,上海提出建设“信息港”的战略目标,并率先建设“城市光网”。10 多年来,上海信息基础设施能级快速提升,普遍接入能力显著加强,基本实现信息通信服务在市域范围内的按需接入;同时,国际通信能力实现跨越式增长,省际通信能力继续保持国内领先。目前,在上海登陆的国际海底光缆总数已达 6 个系统、10 条海缆,系统总容量超过 4 Tb/s,互联网国际出口带宽突破 300 Gb/s;互联网省际出口大幅提升,2009 年底已成为全国首个达到 T 级规模的城市,2010 年达到近 1.5 Tb/s;城市光网光纤到户,用户接入带宽能力最高已达 100 M,2009 年覆盖 75 万户。截至 2010 年底,全市互联网用户普及率超过 68%,移动电话普及率达 122%,家庭宽带用户达到 440 万户,有线电视数字化整体转换完成近 230 万户,IPTV 用户突破 100 万,并初步实现 3G+Wi-Fi 无线宽带通信网络覆盖;新技术在网络中的应用得到不断深化,以软交换、3G、IP 多媒体子系统(IMS)、基于光纤的多种接入方式(FTTx)等为代表的下一代网络技术得到规模应用,下一代 IP 协议(IPv6)、移动通信长期演进(LTE)、NGB 等新技术的应用部署取得突破;超算中心高性能计

算系统总计算能力达到 230 万亿次/s,主机资源和服务能力保持全国领先,应用领域持续扩展,服务范围覆盖全国 27 个省市;功能服务设施不断丰富,服务能力有效提升。互联网数据中心(IDC)和外包呼叫中心实现规模发展,运营服务模式不断深化,IT 服务外包模式初步形成。

此外,上海相关部门早在 2000 年就预见到 RFID 技术与物联网未来的广阔应用前景并着手布局。2003 年迄今持续支持 RFID 技术攻关和应用示范,累计资助金额超过 1.7 亿元,其技术已在危险品监控、食品追溯、世博票务、大型赛事及有关行业中大量应用。2010 年 4 月 27 日,上海市政府公布了《上海推进物联网产业发展行动方案(2010—2012 年)》,在技术研发和产业化等各方面推进物联网发展。其中提到将在上海建设涵盖交通、医疗、安防、物流等各领域 10 个物联网应用示范工程。上海强大的经济实力、完善的基础设施使得上海在物联网建设方面与其他城市相比具有得天独厚的优势。目前,上海已率先建成公共物联网统一接入管理平台,对全市的物联网发展起到积极推动作用。

2010 年在上海举办的世博会以“城市,让生活更美好”为主题,将全球智慧城市最新信息科技率先应用于世博园的安防、管理、服务、交通等各个环节,使世博园成为智慧城市的样板。“十二五”期间,上海进一步延伸世博后续效应,明确了智慧城市建设的 40 字指导思想,即“加大投入、强化基础”,“融合发展、提升效率”,“创新引领、自主发展”,“完善制度、保障安全”,“城乡一体、普遍受益”。明确建设智慧城市基本框架的 4 个主要的关注点^[1]:

一是关注信息基础设施能级提升。为适应高速、智能、融合的趋势,上海着力打造“城市光网”,提升信息网络带宽和接入能力;发展 3G、Wi-Fi 等多种技术的无线宽带网,扩大其在全市域的覆盖;推动智能技术、云计算和物联网等新技术的研发应用,加快“三网融合”进程。

二是关注信息技术的广泛应用。加快信息技术在金融、航运、商贸等服务业领域的深化应用,发挥信息化在改造传统产业和激发新兴产业中的作用;围绕城市规划管理、交通综合信息服务、城市应急联动,建设信息化综合管理平台;引导和发挥社会组织开展信息化的积极性,继续缩小城乡之间和不同人群之间的“数字差距”;促进政务信息共享和业务系统的建设,提升政府信息化服务水平。

三是关注信息技术创新和产业化。一方面借助信息技术创新,带动应用模式创新,促进业务形态创新,进而实现产业形态和结构的更新,催生新的信息服务业;另一方面由信息技术创新激发组织机制和管理模式创新,促进企业创新发展,实现企业做大做强。

四是关注信息化的发展环境。继续深化信息安全保障、信息化政策法规体系、信息化人才培养、信息化合作交流等方面的工作,为信息化的新一轮发展提供支撑。

2011 年 1 月 16 日,当时的上海市市长韩正在政府工作报告中指出:上海要加快建设城市光纤宽带网,实现百兆宽带接入能力覆盖 300 万户家庭;加快三网融合,新建覆盖 100 万有线电视用户的下一代广播电视网络系统;加快建设新亚太海底光缆系统,海底光缆国际通信容量继续保持全国 50%以上。当年 9 月,上海市出台了《上海市推进智慧城市建设

2011—2013 年行动计划》，提出加快建设以数字化、网络化、智能化为主要特征的智慧城市，主要内容包括：

- ① 宽带城市、无线城市基本建成。
- ② 信息感知和智能应用效能初步显现。
- ③ 新一代信息技术产业成为智慧城市发展的有力支撑。
- ④ 信息安全总体实现可信、可靠、可控。

通过政府规划引导，推动相关企业重点实施宽带城市、无线城市、通信枢纽、三网融合、功能设施 5 个专项。在新一代信息技术产业中重点推进城市建设管理、城市运行安全、智能交通、社会事业与公共服务、电子政务、信息资源开发利用等 8 个专项。在信息感知和智能应用方面重点推进云计算、物联网、TD-LTE、高端软件、集成电路、下一代网络(NGN)、车联网、信息服务 8 个专项(见表 1-2)^[14]。

表 1-2 《上海市推进智慧城市建设 2011—2013 年行动计划》重点领域

重点专项内容	新一代信息技术产业专项	信息感知和智能应用专项
宽带城市建设	云计算	城市建设管理
无线城市建设	物联网	城市运行安全
通信枢纽建设	TD-LTE	智能交通
三网融合试点	高端软件	社会事业与公共服务
功能设施建设	集成电路	电子政务
	下一代网络(NGN)	信息资源开发利用
	车联网	“四个中心”建设
	信息服务	“两化”深度融合

值得指出的是，对照建设智慧城市的目标，上海信息化建设目前还面临四大发展瓶颈：

一是信息通信服务能力虽然大幅提升，但仍难以满足经济社会发展中快速增长的信息通信需求，特别是在宽带接入速率、服务能力方面与世界发达城市相比尚有较大差距。

二是信息通信技术创新步伐明显加快，但是有效的商业模式、推广机制、配套制度相对滞后，信息化新技术应用和产业化进程尚需加快。

三是集约化的电子政务等网络基础设施、应用集成平台、信息共享体系初步具备，但资源利用效率、整体应用效能和协同应用水平有待提高。

四是有关智慧城市建设整体规划、基础支撑、实践领域和联动机制等尚需统筹思考与分步落实。

1.3 智慧城市的未来趋势

城市是人类群居生活的高级形式,也是人类文明的主要组成部分,现代工业文明与古代农耕文明的最大区别就在于城市的诞生与发育。到了现代社会,城市被赋予了前所未有的经济、政治和科技功能,在一国经济社会发展中发挥着主导作用。有研究预计,随着城镇化进程,到2050年,全球人口中会有一半以上都生活在城市。同时,城市也面临着诸如环境、交通、能源、水资源、住房、失业、卫生、防灾、教育、治安、医疗护理等各方面的挑战和衍生出的城市病问题。为有效解决这些问题,促进城市实现可持续发展,为人类创造更良好的城市空间,近年来各国政府对智慧城市的推展不遗余力,希望通过智慧城市建设,为居民提供更便利的服务。

根据城市自然禀赋的不同,智慧城市的服务内容包括:智慧基础设施、智慧平台、智慧人才、智慧治理、智慧环境、智慧社会、智慧民生、智慧经济八类与民众生活息息相关的智慧服务。

关于智慧城市的发展趋势,不同学者根据其观察的角度不同有不同的结论。中国工程院许庆瑞院士提出了未来智慧城市的金字塔-星状结构模型,认为智慧城市未来的发展趋势是实现“经济—社会—生态”的全面可持续发展,最终满足居民生活的安全感和幸福感。日经BP社CleanTech研究所《全球智慧城市研究报告》提出三大趋势:委托民间民营企业进行分工,高度提升效率的行政服务;为医疗、健康与自然共生等方向,提供舒适居住的环境;提供社群层级而非个人化的服务。IBM《智慧城市白皮书》认为,21世纪的智慧城市,能够充分运用信息和通信技术手段感测、分析、整合城市运行核心系统的各项关键信息,从而对于包括民生、环保、公共安全、城市服务、工商业活动在内的各种需求作出智能的响应,为人类创造更美好的城市生活。中国通信学会智慧城市论坛《智慧城市白皮书(2012)》认为,智慧城市的愿景是使城市“发展更科学,管理更高效,生活更美好”。

(1) 发展更科学。智慧城市建设的首要目标就是要调整城市产业结构、转变城市发展方式,并以此作为带动城市经济发展的主要抓手。通过智慧城市建设,大力发展新一代信息产业,降低产业能耗,使城市的产业结构更加合理,发展更加持续。

(2) 管理更高效。城市畅通高效的运行,必须依靠可感知、可控制、高智能的管理手段。通过物联网、空间地理等技术,对湖泊、森林、大气等自然环境,道路、地面设施、地下管网等城市基础设施,汽车、食品、药品等流动的物品进行实时的管理控制,对本地人口、常住人口、流动人口等进行实时的管理和服务,以实现智慧的交通、智慧的环保、智慧的应急、智慧的基础设施、智慧的食品药品监管、智慧的政务,达到新阶段的人类文明:低碳社会、绿色生态、山水宜居。

(3) 生活更美好。通过高度发达的基础环境和信息环境,市民在任何时间、任何地点、任何渠道都可以享受便捷廉价的社会服务、就业安居、公共交通、医疗卫生、文化教育等融合的公共服务,生活品质获得大幅提升;强大的服务能力让市民感受到归属与认同、便捷与关怀、交流与信任、尊重与成长。物理位置不再成为障碍,人人比邻而居,成为城市家园中日日相伴的朋友,不论是本土居民,还是外来居民,都为工作和生活在一座智能、亲切的城市而感到舒心和骄傲。

中国软件评测中心副主任张少彤认为,智慧城市在智慧准备与智慧应用的助推下,呈现五大发展趋势:

① 完备、高速、泛在的智慧准备将实现人与人、物与物、人与物的充分互联。

② 安全、统一、开放的应用平台将融入社会服务管理的方方面面。

③ 均等、便捷、个性的社会服务将无所不在。

④ 透明、精细、科学的城市管理将助力城市智慧运行。

⑤ 绿色、创新、融合的智慧产业将促进城市经济蓬勃发展。

综上所述,本书认为智慧城市的未来趋势将表现在以下 5 个方面:绿色宜居、平安有序、幸福和谐、集约高效和科学决策。

1.3.1 绿色宜居

所谓绿色宜居有两层含义,一层是绿色,一层是宜居。绿色是手段,宜居是目的。这里的绿色又有两层含义,一是指狭义的绿化与绿化面积,另一个是广义的绿色,即环保、节能、无害等。从自然的绿色到经济的绿色到社会的绿色再到人心的绿色,四者的总和成为推动绿色智慧城市的基本思考。整个地球上的高等生物都是依靠自然界的碳水化合物来供养,自然绿色是绝对不可缺少的。如 2006 年南京提出要东部绿色中心,成为最适宜居住的城市,列出以下指标:城市生活污水集中处理率已达 63.8%,城市空气质量良好以上的天数突破 300 天,小康环境质量综合指数达 74.5,林地面积 12 万 hm^2 ,森林覆盖率 21%,居住区绿地率 39.5%、绿化覆盖率 44%、人均公共绿地 12 m^2 。经济的绿色就是学习自然生态系统,如循环经济。自然生态系统是全循环,是几乎没有任何废弃物的完整高效的循环系统。食草动物、食肉动物(包括人),它们的排泄物排到自然界,分解者将其分解后变成养分供养新的生产者。但是到了人类社会特别是工业社会以后,生产者和消费者越来越庞大,却没有相应的分解者,环保局就成为帮助分解者,但实际是打破了循环。所以说,经济要学习自然的绿色,完成高效循环,尽量减少废弃物和污染物。

20 世纪 90 年代,联合国第二届人类居住地大会提出了“城市应当是适宜居住的人类居住地”的概念,并形成了国际共识,宜居城市成为了 21 世纪新的城市观。绿色宜居城市是城市的发展方向,也是衡量城市综合实力的重要标准。绿色宜居城市建设包括经济发展、城市生产力布局、改善人们的居住环境等相关因素。城市绿色宜居核心是生态文明,通过绿

色环境保护和培育,扩大城市绿色空间,提高全民环境意识,选择绿色生活方式,实现可持续发展。

谈到城市的绿色宜居,需要介绍一个重要概念,即绿色基础设施^[15]。绿色基础设施是近几年来西方国家就生态保护和城市建设等提出的概念,是建立在生态理论的基础上,针对公路、燃气管网等“灰色基础设施”和医院学校等“社会基础设施”等建筑设施概念而提出的,他将城市开敞空间、森林、公园和其他自然地域形成的绿色网络看作是支撑城市和社区发展的另一种必要的基础设施。这些基础设施维持着空气和水质的清洁,保护自然资源,提升人们的生活环境与质量^[2]。

绿色基础设施比较准确的定义是指具有内部连接性的自然区域及开放空间的网络,以及可能附带的工程设施,这一网络具有自然生态体系功能和价值,为人类和野生动物提供自然场所,如作为栖息地、水源、迁徙通道等,它们的总体构成保证环境、社会与经济可持续发展的生态框架^[14]。在比较微观具体的语境中,可以指代绿地斑块、廊道或者具体的相关工程设施,如洪泛控制体系、水资源净化设施,或者一片次生林,甚至一棵树、屋顶绿化等,都可以被称为绿色基础设施。绿色基础设施包含各种天然和得到恢复的生态系统和景观要素,它们构成一个既有“网络中心”又有“链接环节”的网络系统。

浙江大学科教发展战略研究中心陈劲教授认为^[16],绿色宜居包括:自然舒适的健康环境、新型社会形态的城市网络和绿色智能的生活社区。依托智慧技术为市民的衣、食、住、行等生活和工作的各个环节提供便捷的服务,提高一切资源的利用效率;可以通过对信息的综合分析,为人类的未来活动或者突发事件提供指导、参考。绿色宜居对环境的要求:一是清洁的空气和卫生健康的水流;二是有宁静的社区、美丽的公园和较多的绿化面积;三是建筑和周围的自然环境协调、和谐;四是具有持续性,注重公平地满足现代与后代在发展和环境方面的需要。新型社会形态的城市网络,城市代表着新的社会形态,表现为城市设计、城市生态、基础设施、生产设施的网络化的一些特征。绿色智能的生活社区。社区将成为未来绿色智慧城市中市民生活和工作的基本单元。由于智能技术和商业模式的支持,城市的物质、能量、信息可以被高效利用,城市生活安全、舒适;网络营销、物流等新兴产业的兴起,城市市民的生活和工作区域逐步交融,人流、物流、信息流通畅、增效,绿色智能社区成为市民生产和生活的聚集地和人类理想的家园。陈劲教授还列出了智慧城市绿色宜居的部分行业构成:

① 大气污染防治:包括城市灰霾治理、电除尘、袋式除尘、脱硫、工业有机废气净化、室内空气净化、机动车排放污染治理、脱硝、二噁英控制、高效燃烧等。

② 节水与再生水利用:包括城市水务、城市污水处理、工业废水污染控制与工业节水、饮用水安全、水污染治理与资源化等。

③ 城市固体废弃物无害化资源化利用:包括城市生活垃圾处理、工业固体废物处理技术与装备。

④ 节能智能控制:包括电力系统智能控制、工业系统自动化控制(DCS、PLC)、楼宇智

能控制、交通系统智能控制、应急电源,以及相关的研究、产品开发和商业化应用。

⑤ 清洁技术新材料:包括新型储能电池技术和墙体保温隔热材料。

⑥ 绿色照明:包括半导体材料、芯片,半导体照明封装、半导体照明应用等。

⑦ 新能源:包括太阳能、风能、水电、生物质能、热泵技术、洁净煤技术等。

⑧ 咨询服务:包括技术咨询、规划、设计、投融资等。

⑨ 城市公用事业服务:包括食品安全、物联网等。

⑩ 信息技术服务:包括软件外包、IT 系统服务、城市一卡通、机电系统服务。

下面介绍法兰克福的绿色宜居案例。法兰克福的绿色宜居,体现在以下几个方面:环城绿带(Green Belt)、被动式住房、控制大气排放、垃圾再利用、水资源管理。

(1) 环城绿带。法兰克福的智慧城市建设主要是由法兰克福环保局负责。与其他城市相比,法兰克福更加注重绿色发展,其目标是建设绿色城市。法兰克福全市绿化覆盖率达 52%,由花园、公园、树林、水泽和沙丘等多样化地貌组成,人均占有公园绿地达到 40 m²。法兰克福最大的绿化动作当属持续 20 多年之久的建设环城绿带计划,目前,长达 75 km 的城市外围“法兰克福绿化带”基本建成,不仅成为城市绿化屏障,还通过数次立法,征求市民意见,增添了许多休闲娱乐设施,并向市民开放。

(2) 被动式住房。法兰克福提出了以住宅建筑节能为核心的“节能家庭方案”。这种被动式节能建筑特点体现在两个方面:一是注重房屋的保温密闭性,二是充分利用可再生能源。

(3) 控制大气排放。在环保方面,法兰克福除了关注绿化以外,还比较重视空气的质量。特别是在控制二氧化碳排放方面,采取了大量的应对措施:低排放公交车,天然气汽车,低排放区域,鼓励自行车出行,电动交通。

(4) 垃圾再利用。随着相关技术的成熟,除了常规的填埋和焚烧外,法兰克福正在越来越多地利用生物技术降解垃圾,将之转化为电能和热能。如利用生物发酵剂处理厨房垃圾,将法兰克福全市每天 1 000 t 以上的生活垃圾转化为无公害生物有机肥,可利用这些有机肥培植无公害蔬菜及花卉等。目前,法兰克福有 30 000 户家庭正在使用垃圾发电厂生产的电力,这在一定程度上也能够节约诸如油、气、煤等石化燃料并降低二氧化碳排放,从而保护环境。

(5) 水资源管理。虽然德国整体而言水资源丰富,但政府也不忘通过各种手段鼓励节约利用水资源。如鼓励和帮助居民购买雨水收集设备,主要是通过房顶收集雨水,雨水经过管道和过滤装置进入蓄水箱或蓄水池,再通过压力装置把水抽到卫生间或花园里使用。此外,保护水资源免受污染和提高水质也是政府环境保护的重要组成部分。为了保障水源安全,在含水层周围按不同的距离划分了三级水源保护地带,其中在采水点周围 10 m 范围内的一级保护带要求最为严格,禁止一切有污染的物质渗入地面,违者将被罚以巨款。为了保证自来水质量,从生产、传输和监管等各环节都作出了详细的规定,目前德国自来水都达到了直接饮用的水平,而法兰克福对水质的要求更高,目前已经达到每小时一测的程度。

1.3.2 平安有序

城市平安是对城市运行的最低要求。一座城市文明程度高低最重要的指标之一就是这座城市的秩序如何。所谓平安有序,“平安”是指通过智慧城市的智能化信息技术,如安防网络等,来保证整个城市居民的生活场所安全。城市平安不仅仅是社会治安一项内容,还包括城市交通状况和城市消防服务,以及各种人为灾害(包括战争、恐怖袭击、威胁城市安全的重大火灾、环境污染等)和自然灾害的预警和处警等内容。“有序”是指城市运行更有秩序,包括城市软硬件系统的有序和居民行为的有序。城市有序包括市容美观、市场有序、交通畅通、素质提升等。城市的公共秩序,是检验市民文明素质高低的一面镜子。闯红灯、不走斑马线,驾车乱抢道、办事脸难看都是城市无序的表现。

智慧城市建设为平安有序提供了支撑。运用物联网、云计算等新一代信息技术,解决智慧城市建设和运营中所遇到的安全问题,通过构建安全保障体系,确保在充分互联、资源整合和数据融合背景下的智慧城市的安全可靠,提高智慧城市的安全保障能力。

未来的智慧城市应有其浓厚而独有的城市底蕴,良好的秩序如社会治安稳定、交通文明有序等是每座美丽城市共有和必不可少的。有位游客描述,有一次在巴黎街头,他们准备过马路时,这时有人发现黄色信号灯在闪烁,走在前面的人在同伴的提醒下,又退了回来。可是,等候直行的第一辆汽车驾驶员——一位老太太却带着笑脸,作出了一个“请通行”的手势,全没有国内汽车“抢灯通过”的架势,也没听见其后面车辆不停鸣笛的催促声。其情景让他们感动了很久,也感慨颇多。巴黎也因此在他们心中美丽了不少。城市平安有序,例如交通有序,需要有严格的规则,更要有一群自觉遵守规则的文明人,包括遵守交通规则的行路人和有礼让风范的开车人,而我们的现状与国际大都市确实有一定的距离。

1.3.3 幸福和谐

人类社会发展到今天,从根本上来讲,仍是要解决人与自然的可持续发展、人与人的协调发展以及人对自我的认知与幸福感提升。而智慧城市,就是用现代化科技手段去更好地协调处理好这三个关系。智慧城市的目标是幸福指数最大化,提高市民的生活幸福感。幸福城市含了GDP以外的东西,包括精神方面的度量。市民的生活幸福感包括5个方面的外延,即城市环境完备智能、城市经济生动创新、城市功能高效灵敏、城市市民现代幸福、城市管理精准高效。市民的生活幸福是由城市的智能化做支撑,包括感知泛在信息基础设施、城市智能运转、经济智能创新运营和政府智能管理等方面。

21世纪,人类面向新的文明,智能经济、智能城市使新的文明也会形成一个新的发展浪潮。不仅仅是智能经济、智能城市,它也会迎来其他文明的发展浪潮,一是生态农业、循环工业和持续服务产业为特征的绿色经济浪潮,实现人类发展与生态环境的和谐、持续。二

是生态环境、绿色经济和政府善治为特征的幸福社会建设浪潮,实现人类发展与社会环境的和谐、持续。第三是以智能管理、智能经济和智能社会为基本内容的智能城市建设浪潮,人类发展将走向自由。

智慧城市的理念更强调城市可持续发展和市民幸福度的提高。以北京市石景山区为例,环卫工程车辆通过安装传感器,智能感知道路卫生状况和车辆油耗,自使用以来,道路保洁率提高了8%,燃油节约率达到15%。而停车引导等在交通领域的应用,不仅让市民出行更加方便,还为缓解交通拥堵、减少拥堵产生的汽车尾气排放发挥了重要作用。

在“中关村国家自主创新示范区”的智能体验中心,以环保节能为主线的体验式展示让体验者大开眼界。体验者可以从演示中感受到特大城市智能交通系统的畅通,如不停车电子收费、智能停车引导、交通诱导与信息服务等。智能体验中心作为新的载体,能够通过现代化多媒体技术对感知城市的发展理念、建设愿景进行完整诠释,可以为感知城市理念的宣传起到积极作用。

数据显示,截至2011年5月,我国在建的智慧城市数量已达36个,密布在环渤海、长三角、珠三角地区,三大区域的感知城市数量占了总数的74%。其中,环渤海地区有9个,占总数的26%;长三角地区有11个,占总数的31%;珠三角地区有6个,占总数的17%。此外,武汉城市群、成渝经济圈、关一天水经济圈等中西部地区也已经呈现出较好的发展态势。从现在各地建设感知城市的情况来看,建设目标主要围绕智能、惠民、绿色、畅通、幸福等方面来设定。而实现这些目标的重要前提是社会主体即政府、企业、公众对感知城市的认同,认同其为交通、环境、生活等各方面带来的变革和影响。

1.3.4 集约高效

李德毅院士认为,智慧城市建设的5大要素是安全、高效、有序和谐、绿色和智慧。当前我国城镇化进程进入加速发展时期,中国的城镇化问题和高科技已成为影响21世纪最重要的两大主题,城市发展中各项建设所需资源的迅速扩张,加上社会对资源的配置方式和居民对资源的使用方式粗放,造成资源急剧减少(如土地资源)、资源严重浪费和环境污染。在城市资源越来越紧缺的今天,为了节约资源、保护环境,许多城市开始对其中心区建设进行深度思考和改造实践,希望寻找到更集约、更高效的城市发展模式,城市集约化是未来的一个发展趋势和方向。以土地资源为例,我国国土幅员辽阔,但可用土地资源短缺;同时,我国城市化经济增长方式面临由粗放型向集约型的根本性转变。所以,无论从保护土地资源、城市发展和城市土地合理配置的角度,还是从经济增长方式最佳化的角度,城市土地集约利用的论题都显得十分重要^[19]。所谓集约高效,是指城市资源配置集约、经济发展方式低碳高效、公共管理和服务高效、民生改善高效、对资源和环境的使用更加科学等。集约高效的智慧城市包括城市的建设和运营两个方面。

集约高效的智慧城市建设不是另起炉灶,不是大量开工建设新的信息化应用系统,必

须充分利用和整合已有的各级各类资源,着力打破固有的行业 and 部门壁垒,加强行业之间、部门之间的资源整合力度,促进共享、力求集约。围绕城市运行数据的采集、加工、提炼、分析、应用等开展的融合工作是智慧城市建设的核心与关键环节,要大力推进城市范围内数据融合,通过数据融合真正实现信息与智慧决策的融合,从根本上促进实现政府部门的协同共享、行业的互动协调、城市的精细化运行管理、人与自然的和谐相处。

集约高效智慧城市建设需要市民积极参与,要把关注民生、保障民生、改善民生作为智慧城市建设的出发点和落脚点,着力解决就业、社会保障、医疗卫生、教育、食品药品安全等人民群众最关心、最直接、最现实的利益问题,让老百姓真正体验到智慧城市建设带来的生活品质改善,提升人民的幸福指数。要在应用体系的规划设计和产业发展上注重资源节约和环境友好,形成一种节约、高效、低碳的智慧城市建设模式。

智慧城市对集约高效的支撑,可通过加强就业、医疗、文化、安居等专业性应用建设,通过提升城市建设和管理的规范化、精准化和智能化水平,有效促进城市公共资源在城市范围共享,积极推动城市人流、物流、信息流、资金流的协调高效运行,提升城市运行效率和公共服务水平。

集约高效的智慧城市价值实现是智慧城市整体综合功能满足人与自然健康发展的程度,最终目标是实现城市资源能源清洁高效、自然环境健康宜人、基础设施完善舒适、社会环境和谐文明。

1.3.5 科学决策

根据国外相关学者的观点,智慧城市为科学决策提供了支撑,智慧城市在其运行中将信息通信技术应用到城市运营和发展的方方面面,通过计算机化的系统结构有效完成感知、处理、决策等复杂行为,从而使得城市环境更加宜居、城市运营更加智能、城市管理更加高效^[11]。通过智慧产品、智慧技术、智慧决策系统和智慧创新系统,促进城市经济健康运行、社会和谐发展、生态绿色宜居,即智慧助推服务、助推发展。

智慧的城市如同智慧的人,应当在具备完善的基础设施的前提下,广泛地搜集社会各方面的信息、资源,进行分析处理,从而形成各类完备的数据库。在完善的数据资源基础上,通过运用物联网、云计算、大数据等先进的信息技术,建设各类应用平台系统,实现对经济社会各方面事务的有效预测和监控,协助领导者作出科学的决策,提高城市的管理水平,从而为公众提供更加智慧的服务。

社会服务可以通过智能手机、电脑等终端多渠道获取,并且支持个人定制等个性化服务;城市管理的关键领域实现信息化转变,协助领导科学决策。

透明、精细、科学的城市管理将助力城市智慧运行,透明、精细的城市管理是行政管理体制改革创新的方向。智慧城市在完善的数据库和应用平台的支撑下,城市资源将实现共享共用,部门之间将会有更多的协同办公,从而提高城市管理的运行效率;在各种数据报

表、智能检测系统的支撑下,管理者的决策将更加科学。

智慧城市的技术支撑,为公共决策的公众参与提供了可能。例如,在智慧城市的智慧医疗方面,利用基于区域卫生信息平台,可为居民、医药卫生机构、卫生管理机构提供安全完整的个人健康、卫生服务、综合管理等信息,为实现创新型的业务协同和领导宏观决策提供技术与数据支撑。再如,欧盟城市的市民、各个利益相关方通过参与决策、倾听决策等多种多样的形式参与智慧城市的建设。鼓励对日常可能发生的突发事件所涉及的跨政府部门、跨欧盟区域的政府在线合作,以满足各行业、各类组织和公众的需求,提高城市政府的管理能力。社区居民也能够自主发起一些社区智能化项目,在项目实施过程中居民的参与将成为一种社会发展趋势。

◇参◇考◇文◇献◇

- [1] 王家耀,刘嵘,成毅,等. 让城市更智慧[J]. 测绘科学技术学报,2011,28(2): 80-83.
- [2] 杨静,潘国锋. 建设城市绿色基础设施,打造绿色宜居城市[C]. 2011 城市发展与规划会议论文: 296-298.
- [3] 张永民. 智慧城市总体方案[J]. 中国信息界,2011(3): 12-21.
- [4] 王德恒. 关于建设智慧城市的战略思考[J]. 城市建设理论研究(电子版),2012.
- [5] 许庆瑞,吴志岩,陈力田. 智慧城市的愿景与架构[J]. 管理工程学报,2012(4): 1-7.
- [6] 方晓华. 浅谈智慧城市的构建[J]. 华人时刊(中旬刊),2012.
- [7] 赵伟娟,等. 浅谈智慧城市的建设和发展[J]. 城市建设理论研究,2013.
- [8] 宋刚,等. 创新 2.0 视野下的智慧城市[J]. 城市发展研究,2012.
- [9] 陈立,李春香,李志勇. 浅议智慧城市的“躯体、经络与大脑”[J]. 计算机光盘软件与应用,2012(8): 79.
- [10] 张少彤,王芳,王理达. 智慧城市的发展特点与趋势[J]. 电子政务,2013(4): 2-9.
- [11] 陈柳钦. 智慧城市: 全球城市发展新热点[J]. 青岛科技大学学报(社会科学版),2011.
- [12] 原松华. 智慧城市: 探寻未来城市发展之路[J]. 中国发展观察,2013.
- [13] 巫细波,等. 智慧城市理念与未来城市发展[J]. 城市发展研究,2010.
- [14] 徐玉春. 关于推进“智慧天津”战略及对策研究[D]. 天津: 天津大学,2012.
- [15] 刘晓丽. 地理信息泛在服务若干关键技术研究[D]. 青岛: 山东科技大学,2012.
- [16] 李开然. 绿色基础设施: 概念,理论及实践[J]. 中国园林,2009.
- [17] 许伟. 城市土地集约化利用及其评价研究[D]. 重庆: 重庆大学,2004.

第2章

智慧城市的技术体系总体框架

智慧城市是指通过物联网、云计算、大数据、人工智能等技术，实现城市基础设施、公共服务、城市管理等方面的智能化、网络化、协同化。本章主要介绍智慧城市技术体系总体框架，包括感知层、网络层、平台层、应用层四个层次。

(1) 感知层：负责采集城市运行数据，包括交通、环境、公共安全、民生服务等。感知设备包括摄像头、传感器、RFID标签等。

(2) 网络层：负责数据传输和交换，包括有线网络、无线网络、移动网络等。

(3) 平台层：负责数据整合、分析和处理，包括大数据平台、云计算平台、物联网平台等。

(4) 应用层：负责提供智慧城市应用，包括智能交通、智能安防、智能环保、智能民生等。

本章主要讨论智慧城市技术要求、智慧城市参考模型和智慧城市技术体系架构三方面内容。其中,智慧城市技术要求智慧城市建设具有创新性、实用性、智能性、扩展性、兼容性、安全性和规范性;智慧城市参考模型包括概念模型和功能模型;智慧城市技术体系架构包括基础服务层、平台服务层、应用服务层、Gateway 接口、SOA 整合平台和 SOA 平台构架等方面。

2.1 智慧城市技术要求

2.1.1 创新性

目前智慧城市正处在发展的关键时期,也面临着战略方向的选择。当前世界各国,尤其是欧、美、日、韩等发达地区和国家,都在积极开展各项研究与应用,建立相应的城市试点,研发城市智慧应用,以期占领智慧城市的技术制高点。由于目前智慧城市相关的技术及应用还没有公认的解决方案和标准,尚未形成新的技术和市场壁垒,我国与发达国家在这个领域中基本处于同一起跑线上,智慧城市研究是一个从“跟随”到“引领”的重要发展契机,应该把握这个历史机遇,占据智慧城市相关各项前沿技术的领导地位,引领国际相关领域发展方向。因此,在智慧城市的建设中,应注重加强对技术创新性的要求。

(1) 通过大力推进先进技术的研发,建立起领先的智慧城市技术概念体系,引导智慧城市相关部门确立技术创新的战略,推进技术创新投入,增强技术创新能力。

(2) 加强创新体系建设,激发技术创新的内在动力,提升技术创新能力,着重突破智慧城市的各项关键技术,形成一系列具有自主知识产权的技术、产品和专利。

(3) 大力推动智慧城市相关各种新兴产业的发展和新技术应用,打造创新型城市智能服务,创建以人为本的智慧城市应用模式。

(4) 在创新机制上有所突破,深化智慧城市所涉及的多个应用领域改革,加大产学研的紧密结合,在智慧城市建设中发挥政府主导力量,充分调动技术创新的积极性,实现智慧城市的科学合理可持续发展。

2.1.2 实用性

智慧城市的建设应该面向城市居民的基本需求,利用先进的技术手段,以信息化的方

式服务于城市的建设和发展需要。因此,在智慧城市建设过程中,要始终强调实用性原则,以满足城市可持续发展为依据,寻找符合城市特点的应用领域进行建设,完善发展规划,确保智慧城市建设符合城市发展的总体目标。

(1) 坚持以人为本,把智慧城市建设与提高城市居民生活水平紧密结合起来,通过智能的城市服务提高城市居民生活的便利,满足民生需求,构建和谐社会。

(2) 在技术创新和应用上要充分考虑智慧城市各个应用领域的特点,强调智能服务和应用的便利以及可行性,把满足城市智能服务对象作为重要的因素进行考虑,形成实用化的智能服务体系。

(3) 通过智能医疗、智能交通、智能水务、智能物流等专业化系统的建设,提升城市规划和管理水平,促进社会资源的合理分配,推动城市发展向智能化转型。

(4) 通过智慧城市建设过程中新技术的应用,实现相关产业升级,有效地推进物联网、传感网、通信等先进科学技术的应用,推动全社会的科技投入,加快智慧城市相关设备和产品的研究发展,提升城市发展的整体信息化水平,促进信息化和传统行业的融合与发展,为城市产业升级和结构转型提供技术保障。

2.1.3 智能性

智慧城市的建设应面向城市服务功能的智能化,以各种城市资源和采集到的数据为基础,通过智慧城市的技术手段将城市服务的提供者和服务对象有机地连接在一起,为城市智慧化发展提供广泛的技术支持。

(1) 在智慧城市建设中,应大力扩展各类基础数据的快速及时感知和采集手段,提高多样化的数据感知以及互联互通的沟通效率,增强互动整合的数据共享和处理能力,为城市各类应用服务提供可靠的数据服务基础。

(2) 智慧城市的建设应打破传统上按行政、行业划分的服务方式,大力提倡一站式智能服务,积极推进服务水平提升,提高城市服务居民的满意度、企业的竞争力以及政府的执政水平。

2.1.4 扩展性

智慧城市的建设不是一蹴而就的,通常面临着建设周期长、涉及面广、需要根据不同的城市特色进行阶段性建设的挑战。因此智慧城市的可扩展能力是一个核心的要求,在建设过程中必须考虑到技术进步以及未来应用环境和目标的变化,保证智慧城市在建设过程中不断调整和自我完善,满足不断变化的各类需求。

(1) 随着技术水平的不断进步,相关产业不断整合,智慧城市的应用系统需要不断更新升级,以保证应用的高效和实用,确保智能服务的有效扩展。因此智慧城市的建设是一个

动态不断进步的过程,随着应用环境的不断变化和领域的不断扩展以及需求的逐步演化,智慧城市相关的技术和应用系统也需要不断改进扩展,以满足总体建设的要求。

(2) 随着智慧城市建设的深入,智慧城市的参与者也会对服务水平提出越来越高的要求,智慧城市的新需求会不断涌现,因此在建设过程中必须意识到这种变化,不断提高建设的目标,完善服务的内容,提升服务的质量,在相关技术选择和应用上也应该作出相应改进和调整,通过扩展已有的智能服务来满足智慧城市服务对象的新需求。

2.1.5 兼容性

进行智慧城市建设的城市具有不同的发展水平,所涉及的产业也多种多样,建设目标也不尽相同,在建设过程中不可避免地具有技术手段多样、应用水平不一致等困难。因此在建设过程中需要考虑不同城市发展特点对不同技术和行业的特殊要求,在智慧城市发展过程中注重突出不同的城市特色,满足智慧城市差异化的发展需求。

(1) 由于智慧城市建设涉及的应用领域多种多样,技术手段各不相同,因此智慧城市在建设过程中需要重点考虑不同应用领域和技术手段之间的兼容性问题,要统筹考虑,避免建设过程中各自为政。

(2) 目前有一些城市已经推行了一系列智能服务系统,积累了智能服务的丰富经验,智慧城市的建设应广泛借鉴目前已有的城市智能化建设基础,兼容已有的城市智能化系统和技术。但这些既有智能服务系统之间通常存在使用标准不统一、技术手段不一致的问题,因此,在建设智慧城市的过程中,还应大力推进已有系统的融合,提高各个系统和技术接口的标准化程度,安全集成既有系统的数据和服务。

(3) 智慧城市的建设通常根据城市发展水平和特色进行差异化建设,因此,不同城市对于智能服务的需求有共性也有差异性,在智慧城市的建设中,需要对于差异化的建设目标具有兼容的能力。

2.1.6 安全性

随着信息技术的发展和应用,信息安全一直都是信息化过程中最重要的内容之一,在智慧城市的建设中尤为如此。保障数据和服务的安全是智慧城市建设的重要前提。

(1) 智慧城市的建设涉及城市功能的各个方面,需要各种信息化子系统的支持。在实际应用中,应加强智慧城市系统建设的安全性保障,在技术手段上要保证智慧城市系统高效稳定运行的要求,同时加强安全保障规定和措施的推行,完善智慧城市信息化系统的基础建设。

(2) 智慧城市在提供智能化服务的同时,不可避免需要获取和存储城市服务对象的各种信息,其中包含大量隐私性较强的信息,需要进行必要的保护。智慧城市的建设需要加

强对敏感信息保护的关注,从技术手段和管理上保障智慧城市参与者的隐私。

2.1.7 规范性

目前国内许多城市都开展了智慧城市的建设,开发和部署了各种信息化子系统,但普遍存在缺乏统一规划、缺乏相应技术标准和法律规范支持,这种现状严重制约了智慧城市建设,容易造成各系统之间信息交互不畅,形成信息孤岛,因此需要大力推动技术标准和规范的建设。

(1) 在智慧城市的建设中,要加快建立通用的智慧城市技术标准体系,形成国家级智慧城市技术指导规范,同时推动智慧城市技术体系标准的推广和应用,引导智慧城市建设中参与各方采用共同标准,在实际建设过程中不断完善和发展。

(2) 开展广泛的国际合作,大力参与国际各类相关标准的讨论和制定,积极把握技术标准建立的话语权,引领智慧城市的技术发展趋势,推动我国智慧城市技术标准的国际化。

(3) 智慧城市相关的各类技术规范 and 标准应充分借鉴已有的国际标准、国家标准和行业标准,制定的规范和标准应具备前瞻性、科学性、完整性、延续性和可操作性。

2.2 智慧城市参考模型

从城市的主体及其活动出发,定义智慧城市的要素空间。依据要素空间,给出智慧城市的参考模型,包括概念模型、智慧模型、功能参考模型和体系架构参考模型。

2.2.1 智慧城市的概念模型

城市是人类文明的重要组成部分,城市也是伴随人类文明与进步发展起来的,真正意义上的城市是工商业发展的产物。城市是人口较为稠密、工商业较为发达的地区,一般包括住宅区、工业区和商业区等机能分区,并且具备行政管辖功能。城市的行政管辖功能可能涉及较其本身更广泛的区域。城市有不同的等级分类,如小城市、中等城市、大城市、国际化大都市、世界级城市等,对城市能级分类的一个标准是人口的规模。按城市综合经济实力和世界城市发展的历史来看,城市分为集市型、功能型、综合性、城市群等类别,这些类别也是城市发展的各个阶段。

(1) 集市型城市。属于周边农民或手工业者商品交换的集聚地,商业主要由交易市场、商店和旅馆、饭店等配套服务设施所构成。处于集市型阶段的城市在中国称集镇。

(2) 功能型城市。通过自然资源的开发和优势产业的集中,开始发展其特有的工业产

业,从而使城市具有特定的功能。不仅是商品的交换地,同时也是商品的生产地。但城市因产业分工而形成的功能单调,对其他地区和城市经济交流的依赖增强,商业开始由封闭型的城内交易为主转为开放性的城际交易为主,批发贸易业有了很大的发展。这类型城市主要有工业重镇、旅游城市等。

(3) 综合型城市。一些地理位置优越和产业优势明显的城市经济功能趋于综合型,金融、贸易、服务、文化、娱乐等功能得到发展,城市的集聚力日益增强,从而使城市的经济能级大大提高,成为区域性、全国性甚至国际性的经济中心和贸易中心(大都市)。商业由单纯的商品交易向综合服务发展,商业活动也扩展延伸为促进商品流通和满足交易需求的一切活动。这类城市在中国比较典型的有直辖市、省会城市。

(4) 城市群(或都市圈)。城市的经济功能已不再是在一个孤立的都市体现,而是由一个中心城市为核心,同与其保持着密切经济联系的一系列中小城市共同组成的城市群来体现。如美国东岸的“波士顿—华盛顿城市带”、美国西岸的“旧金山—圣迭戈城市带”和美国中西部的“芝加哥—匹兹堡城市带”,日本的东京、大阪、名古屋三大城市圈,中国台湾的“台北—高雄”城市带(台湾岛西部走廊),英国的“伦敦—利物浦”城市带等。广州所在的珠江三角洲地区和上海所在的长江三角洲地区实际上也正在形成两个经济关系密切的珠江三角洲城市群和长江三角洲城市群,整体经济功能日益凸显^[1]。

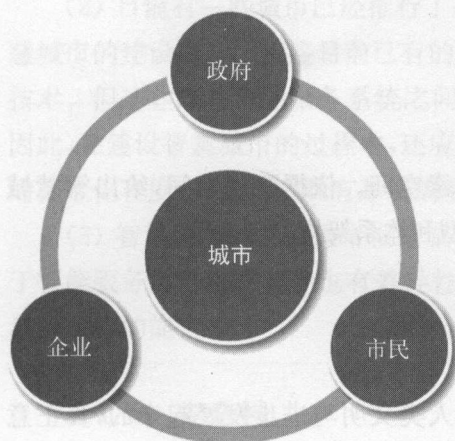


图 2-1 构成城市的主体

无论何种类型的城市,一般而言,市民、政府和企业(包括组织)是构成城市的主体,如图 2-1 所示。市民即城市的居民,广义的政府包括立法机关、行政机关、司法机关和军事机关,狭义的政府仅指行政机关。城市政府,一般是指地方政府。政府作为社会最大的公共组织,不仅负有价值导向、安全保卫、维持秩序的责任,也肩负了政治统治、经济管理、社会协调的社会寄托。政府具有公共属性,政府的职能是为公众服务。

政府职能也涵盖了社会公众生活的全方位,并且随着社会的改变不断发展。政府也同时为公众福利提供方便,创造条件,甚至通过直接提供公共物品提高公众的生活水平。从程序的角度,政府的职能包括了决策、计划、组织、指挥、协调、控制等,其中,决策是政府的核心职能。

政府的职能可以分为政治职能、经济职能和社会职能。其中政治职能是政府最主要的职能,这主要体现在维护现存的政治秩序上。政府的首要任务是维护社会的正常运转,其次就是制定有利于国家与社会的政策,推动社会的进步。为了保证政府职能顺利发挥作用,军队、警察、监狱等暴力机关对外保卫国家主权,对内维护政府统治,打击犯罪。在经济方面,政府通过财政和货币政策调节经济发展,保证宏观经济的稳定增长。经济低迷时期,

通过减免税收、调节税率、加大支出等手段,促进投资和生产;经济过热时,减少政府支出,减少货币供应量,提高利率,抑制需求,降低物价。同时政府通过税收和政府支出等手段,对社会财富进行再分配,以保证社会公平与稳定。为了控制垄断和外部不经济行为,政府通过提高资源配置手段来保证企业间的公平竞争。政府还需要提供充分的政策和资金的支持,积极创造条件,促进经济结构的优化和转型,鼓励新兴产业的成长。政府还提供国防安全、公共工程、社会福利等公共产品为社会公众服务。政府作为社会公共组织,负有维护社会治安和秩序的职责,减少各种社会问题。牵涉到社会公共利益的事务,例如交通、卫生、教育、文化、环境保护等,也往往是政府的责任。

企业是指把人的要素和物的要素结合起来、自主地从事经济活动、具有营利性的经济组织。这一定义的基本含义是:企业是经济组织;企业是人的要素和物的要素的结合;企业具有经营自主权;企业具有营利性。根据实践的需要,可以按照不同的属性对企业进行多种不同的划分。例如:按照企业组织形式的不同,可以分为个人独资企业、合伙企业、公司企业;按照企业法律属性的不同,可以分为法人企业、非法人企业;按照企业所属行业的不同,可以分为工业企业、农业企业、建筑企业、交通运输企业、邮电企业、商业企业、外贸企业等。

城市中主体参与的活动涉及文化、健康、教育、政务、建筑、金融、交通、资源、购物、旅游、生活、食品、环保、公共安全等诸多方面,如图 2-2 所示。

智慧城市主要体现在各种城市活动的智慧性。对城市而言,所谓智慧,就是能够快速响应、高效率 and 高质量地进行城市的各种活动。或者说,智慧城市系统应该使城市主体能够及时得到所需的高质量服务。在各种城市活动中,政府、企业、市民之间通过各种应用系统获取、传递、交换相应的数据与信息。新一代信息技术(Information and Communication Technologies, ICT)的发展为智慧城市的实现创造了可能,通过健全、透

明、充分的信息获取,通畅、广泛、安全的信息共享和有效、规范、科学的信息利用,使城市中的个人、企业、组织、政府、自然系统和人造系统之间更和谐的交互,提高城市运行和管理效率,改善城市公共服务水平,增强处理突发事件的能力,让城市成为和谐社会的中枢。

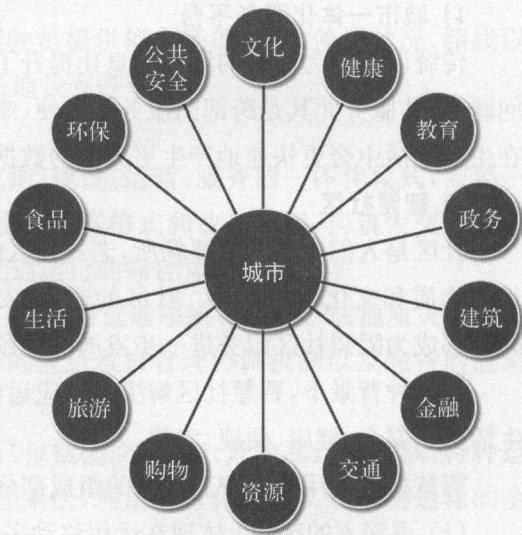


图 2-2 城市中主体参与的活动

2.2.2 智慧城市的功能参考模型

智慧城市整体解决方案不是应用和设施的简单叠加,而是在一体化顶层设计指导下的

共建共享。智慧城市解决方案是需求导向的,这个需求不仅是当前传统手段没能满足的需求,还包括未来发展变化的需求。要更好地满足这些需求,就需要打破部门信息孤岛,进一步共享资源,并实现服务的汇聚、融合和创新。

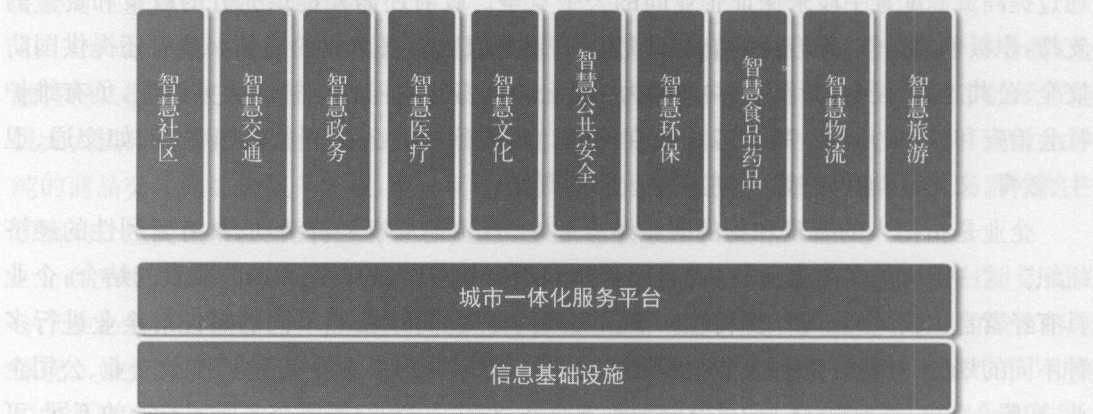


图 2-3 智慧城市的功能模型

1) 城市一体化服务平台

传统以部门为中心的城市信息化提升了各自的业务能力,但也造成了严重的信息孤岛问题,公共服务尤其是跨部门服务不方便、难推进。随着智慧城市建设和网络社会的发展,在生产生活中会更快速地产生更大量的数据,如果没有转变,还会产生更大的信息孤岛。

2) 智慧社区

社区是人们生活的主要场所,老式社区的运营及管理方式已远远不能满足人们日益增长的物质和文化生活需要。社区的通信、安全、环保、社区居民信息的收集、处理及共享等问题都成为阻碍社区服务进一步发展的瓶颈。

在这种背景下,智慧社区解决方案应运而生。智慧社区的基本服务理念是使社区居民生活更加安全、健康、便捷、幸福。

智慧社区应用作为智慧城市的组成部分,与传统的应用建设存在以下区别:

(1) 更深入的感知:体现在运用移动采集设备、3G 网络、物联网技术,实现信息的实时采集,达到及时、准备、双向的感知。

(2) 更深入的互联:体现在利用平台的数据交换和应用集成功能与市区条线系统的整合以及与外部各类基础数据库的互联打破数据壁垒,另外通过统一管理支撑平台和管理平台的建设将区内运行产生的数据也汇聚到信息资源库,这样各类数据在社区层面都获得沉淀,为社区内各类对象提供主动的、便捷的服务,同时为上层各类应用和统计分析提供数据保障。

(3) 更深入的智能:在深入的感知、互联的基础上提供更全面的辅助、分析、增效。为各类智慧应用提供数据、平台功能支撑。

(4) 管理服务创新: 在服务模式、服务手段、服务方法上创新, 形成可持续发展的机制和社区服务商业模式, 根据社区服务的实际发展, 优化社区组织架构, 创新服务模式。

3) 智慧交通

利用先进技术提高大通道等交通基础设施和交通运输工具的使用效率, 推行智能交通系统是增加交通有效供给能力的主要措施。同时通过智能交通系统构建建设、管理、协调、服务一体化的城市交通运营管理机制, 推动综合交通运输体系智慧运行。主要包括以下 3 个方面的需求:

(1) 智慧的出行需求。

① 出行前信息服务: 对于公交时刻表和公交线路、换乘地点、票价、合乘匹配等实时信息, 以及交通事故、道路施工、拥堵路段、个别路段车速、特殊活动安排、气候条件等信息的查询需求。

② 出行中信息服务: 驾驶员通过可变情报板、车内导航设备等收到最优路径、路网交通运行状态息、交通事件、停车场信息、收费站信息、气象信息及路边服务信息等公交方式的出行者通过电子站牌、车载信息屏、手机等方式获得实时的公交运行信息和换乘信息服务。

③ 停车诱导信息服务: 通过多种方式向驾驶员提供停车场的位置、使用状况、路线以及相关道路交通状况等信息, 诱导驾驶员最有效地找到停车场。

(2) 智慧的监管需求。

① 大交通政务管理: 围绕城市智慧交通规划、建设、运营、服务的一体化模式, 需要一方面对现有部分业务系统升级完善, 同时需要规范数据格式和业务流程, 构筑大交通数据中心; 另一方面从一体化管理的角度提供面向交通运行的综合管理功能。

② 应急指挥管理: 建立交通事件自动检测系统, 对交通事故、灾害天气、地质灾害等进行自动监测, 提高交通事件检测效率; 建立完善的应急救援管理协调机制以及完备的应急预案, 实现应急救援工作的统一协调指挥。

③ 营运车辆管理: 建立涵盖各类营运车辆(包括公交、出租、危险品运输车辆、特种运输车辆、普通货运车辆、省际客运等)的运营管理系统, 利用信息化技术实现车辆运输的全方位、全过程的实时监管。

④ 交通监控: 对私家车、营运车辆、船舶等交通运输工具的监测, 主要利用 RFID、传感器和 GPS 等新一代物联网技术, 实现身份识别、运行轨迹描述等监测, 为交通工具的运行、监管和服务提供有效支撑。

(3) 智慧运营需求。

① 公交智能调度与指挥: 建立公交动态运行数据采集系统, 并应用公交动态优化调度系统, 提高公交运营效率和服务水平, 降低公交运营成本。通过完善的公交信息发布系统发布全方位的公交信息, 提高公交服务水平。

② 出租车管理与服务: 建立出租汽车服务管理信息系统, 为建立科学、高效的出租汽

车运行监测机制和行业宏观调控提供决策支持。

③ 维修检测服务：建设车辆综合性能检测信息系统，建立车辆技术管理数据库，使有关管理部门能够及时了解当前车辆综合性能检测情况，一方面实现对道路运输车辆综合性能的监督管理，另一方面实现对综合性能检测站和维修企业的监督管理。

④ 轨道交通运营管理：适应网络化发展要求，借助信息化技术，集成运营生产过程相关专业系统，建成功能齐备、高效安全的运营生产调度平台，提高运营生产管理水平。

4) 智慧政务

智慧政务，指通过深入感知、信息交换、流程整合、智能运算等一系列技术手段和管理手段应用，针对政府业务的全过程环节，提供跨越式和创新式的支撑，为社会公众和企业提供更便利的服务，使得政府工作人员的工作更高效，同时政府业务办理更加阳光透明，整个城市更具吸引力。

智慧政务不仅仅是个概念，而是电子政务发展的必然趋势。相对于传统的电子政务信息化建设，智慧政务主要具备以下几个特点：

(1) 政务信息高度集聚整合：智慧政务要求充分深入地感知政府业务所涉及的各种主体、资源、业务办理过程和结果等各类信息，形成统一的政务信息资源中心，为各种智慧政务应用提供数据支撑。除了传统的自然人、法人、空间地理、宏观经济四大基础库之外，还应包括政府业务办理过程中产生的各类结果信息。

(2) 政务系统灵活按需应变：智慧政务要求政府信息化系统能够灵活响应复杂多变的业务需求，尤其是满足越来越多的跨部门协同应用需求。

(3) 业务办理全程网上流转：智慧政务要求政府内部的各项业务办理按照统一规范实现全程网上流转，包括单部门的内部业务办理以及跨部门的协同业务办理。除了提高政府自身的工作效率，还将大大增加业务办理的透明度，实现阳光政务。

(4) 政府服务多种渠道整合：要由单一渠道服务向多渠道一体化服务转变，在政府服务提供的过程中实现多渠道整合。

(5) 政府管理服务模式创新：基于政务信息高度集聚整合、政务系统灵活按需应变、业务办理全程网上流转、政务服务多种渠道整合，完全可以规划设计一系列创新的智慧政务应用，体现智慧政务的实际效果。

5) 智慧医疗

智慧医疗可以从感知、互联、智能、创新4个方面进行理解：即通过更透彻的感知，全面及时地掌握医药卫生的信息；通过更全面的互联互通，实现跨业务、跨机构、跨行业、跨区域的信息联动与整合；通过更深入的智能化，为用户呈现更加便捷、高效、有价值的信息聚合，而贯穿全过程需要在体制、业务、管理、运营等方面进行创新，保障智慧医疗整体运作体系。

(1) 感知：在现有基础之上，借助物联网技术，将智能传感设备渗透入医药卫生整个生态系统，实现更加透彻、及时的信息获取。

(2) 互联: 在现有政府专网的基础上, 借助 4G、HSPA、NGB 等下一代通信技术, 基于智慧医疗卫生信息平台, 制定互联标准与技术规范, 促进跨业务、跨机构、跨行业、跨区域的信息联动与整合。

(3) 智能: 利用云计算技术, 处理、建模、预测、分析各个环节产生的数据, 深入挖掘信息的内在价值, 并以易于理解的形式予以呈现, 提高工作效率, 提升服务质量。

(4) 创新: 基于医改背景, 结合区域实际情况, 推进体制、业务、管理、运营等方面的创新, 保障智慧医疗整体运作体系。

6) 智慧文化

智慧文化是指通过使用先进的信息采集技术、信息通信技术、信息处理技术和无所不在的宽带基础设施, 满足学习者日益增长的各种需求, 为每个学习者提供个性化的学习服务, 实现重视培养获取运用知识和创造知识的能力的教育体系建设, 逐步构建社区教育体系, 实现虚拟校园, 扩充学校基本功能, 为构建全新的城市大教育体系并为终身教育提供可持续发展的基础。

(1) 建设文化教育行业开放标准规范, 遵循国家教育部及相关行业技术规范指导意见, 制定文化教育行业信息化标准规范, 向行业内部的各类应用提供开放的接口。

(2) 加快教育信息基础设施建设, 建成各级各类学校的数字化教育服务体系, 促进教育内容、教学手段和方法现代化。整合现有资源, 构建先进、高效、实用的数字化教育基础设施。推进数字化校园建设, 以多种方式接入互联网。

(3) 加强优质教育资源开发与应用, 开发网络学习课程, 建立数字图书馆和虚拟实验室。建立开放灵活的教育资源公共服务平台, 促进优质教育资源普及共享。

(4) 逐步构建继续教育、社区教育、老年教育等终身教育体系, 促进和保障全民学习和终身学习的社会, 通过多种信息化手段和方式满足学习者的多种个性化需求, 培养学习者自主学习的状态和习惯。

7) 智慧公共安全

公共安全是关系到整个社会平稳运行和社会中每一个主体生活、工作、发展的要务, 也是构建和谐社会、推动经济发展的必要保障。

智慧公共安全与传统社会管理和突发公共事件应急管理的区别在于其政府管理、社区管理、社会组织管理和企业社会服务的“智慧化”。各个主体之间围绕信息获知、业务联动、管理服务等方面形成城市公共安全管理闭环。

作为社会管理主体的政府、基层管理主体的社区主要肩负着信息发现、信息互通、协调联动、多方管理和领导决策支持的责任, 需要为城市运行的管理者和执行者提供全方位服务。作为社会公共安全其他主体的企业、市民和社会组织, 一方面能够及时获知城市公共安全的相关信息, 另一方面能够作为城市公共安全的参与者在信息发现、业务互动方面尽到相关义务。其中主要包括: 全方位的信息采集和报送、信息流转和处置、信息跟踪和反馈、实现部门信息资源交互和共享、协调指挥与分流命令发布、信息汇聚和辅助分析决策、

对城市运行状况进行监控分析与预警、信息发布、评估管理以及预案制作与管理。

8) 智慧环保

智慧环保是未来环境保护工作发展的方向和目标,此概念是在环境信息化与环境保护业务有机融合的基础上提出的。“融合”是环境信息化的发展之本,即利用更好的、更有效的感知手段提升环境感知能力;用互联互通平台协同环境管理业务,充分交换和共享环境信息资源;在广泛利用信息资源的基础上,提高环境管理效率和科学决策水平。

(1) 智慧的动态感知:利用物联网技术对城市环境质量进行全面的动态感知,对污染源进行全程动态监管,提升环境监管水平,以提高对环境的预测预警能力以及对环境污染事故的应急能力。

(2) 智慧的应用支撑:通过一个智慧的环境信息平台实现系统底层组件、服务的统一部署和管理,为业务系统信息维护、业务处理、流程编排等具体功能的实现提供统一的支持,建立系统开发和运行的基础框架。

(3) 智慧的服务:随着环境管理业务的不断发展,围绕着环境监测、污染控制、生态保护、核安全与辐射管理和环境应急等核心工作,组织开展全方位的环境管理工作。这些工作围绕着共同的监管对象,需要紧密的流程衔接和信息共享,从不断发展的业务应用中再将新的服务作为构件沉淀下来,逐步地丰富环境信息业务构件。

9) 智慧的食品药品

智慧食品药品安全管理指通过物联感知、信息整合、流程整合、智能运算、风险管理等一系列技术手段和管理手段应用,针对食品药品安全监管的全过程环节,构建生产经营者自律、政府监管、社会监督三位一体的食品药品安全管理和服

世界正变得可以感知和度量。通过部署于食品药品传递链的各类传感器,感知身份、感知位置、感知环境、感知状态。世界正变得相互关联互通。宽带、无线、移动通信网络、城市中的各类传感网络,使食品药品传递链中的信息流可以安全地互联起来,使全程管理成为可能,为召回和追溯提供技术支撑。

基于收集数据的相关预测、分析及优化成为可能,促进风险识别和处理,为决策者提供支持,为市民安全消费提供保障。

智慧的食品药品安全管理紧扣和满足各方需求。随着现代社会知识和信息越来越容易获取,对食品药品安全越来越关注。公民需要更加安全、更加健康、更有效率、城市更有吸引力,从而对政府企业的服务创新有更高的要求。

(1) 提高食品药品经营企业信息化水平和风险防控能力:通过先进技术和手段的应用,解决原材料的供应安全问题,企业基于应用自身的身份标识技术,解决假冒威胁,建立品牌。

(2) 实现标准化:从身份编码识别到可追溯性维持,都需要标准化支持。

(3) 监督管理:通过食品药品传递链的信息整合,可提供追溯、召回、防假等一系列应用。

(4) 资源整合和服务整合：对食品药品检验资源和应急处置资源的统一管理，提高资源利用率，缓解资源不足压力。

(5) 知识汇聚：人们需要获取权威的食品药品安全消费知识，通过智慧平台，整合汇聚群体知识和智慧。

10) 智慧物流

智慧物流是指将无线网络、3G 网络、RFID、传感器、云计算等新一代的 IT 技术充分应用到物流行业中，将物联网和互联网整合起来，实现人类社会和物理系统的整合。从而对整合网络内的人员、机器、设备和基础设施实施实时的管理和控制。从而实现现代物流的自动化、网络化、可视化、实时化、跟踪和智能控制。智慧的物流具有以下特点：

(1) 灵活性：能充分利用资源，确保环境可持续发展，但同时还能平衡成本、质量、服务和时间之间的关系。

(2) 可视化：实现整个供应链的可视化，跨价值链的连通，支持智能化的协作。

(3) 信息同步传递：提供对过去、当前和未来趋势的分析，并在供应链中实时传达。

(4) 降低风险：可高效且迅速地发现、降低并调整供应链上的风险。

(5) 以客户为中心：能够满足日益严苛的客户需求，更精确地提供同步供应以及可追溯性。

智慧物流就是要利用新技术和新方法来应对这一挑战。供应链将供应商、业务合作伙伴和客户交织在一个复杂、动态的网络中。他们不仅是经济活动的参与者，不仅仅充当产品和服务的生产者、分销者和消费者的角色。他们共同构成一个复杂的信息网络。智慧物流将促使物理网络和数字网络的融合，将先进的传感器、软件和相关知识整合到系统中。

智慧物流可通过使用强大的分析和模拟引擎来优化从原材料至成品的供应链网络。帮助企业确定生产设备的位置，优化采购地点，也能帮助制定库存分配战略。

11) 智慧旅游

智慧旅游是智慧城市建设的重要组成部分，就是利用云计算、物联网等新技术，通过互联网/移动互联网，借助便携的终端上网设备，主动感知旅游资源、经济、活动和旅游者等方面的信息并及时发布，让人们能够及时了解这些信息，及时安排和调整工作与旅游计划，从而达到对各类旅游信息的智能感知、方便利用的效果，通过便利的手段实现更加优质的服务。智慧旅游有以下 4 个特点：

(1) 全面物联：智能传感设备将旅游景点、文物古迹、城市公共设施物联成网，实时感知旅游产业链上下游运行的核心系统。

(2) 充分整合：实现全市景区、酒店、交通等设施的物联网与互联网系统完全连接和融合，将数据整合为旅游资源核心数据库，提供智慧旅游服务基础设施。

(3) 协同运作：基于智慧旅游服务基础设施，实现旅游产业链上下游各个关键系统和谐高效地协作，达成城市旅游系统运行的最佳状态。

(4) 激励创新：鼓励政府、旅游企业和旅游者在智慧旅游服务基础设施之上进行科

技、业务和商业模式的创新应用,为城市提供源源不断的发展动力。

在信息技术取得突破性进展的背景下,旅游信息化建设正不断向纵深发展。一方面是旅游信息化已经进行到一定的程度,旅游行业各个领域的信息采集与存储具备了相当的规模,通过新技术的应用,积累的信息资源有可能集中发挥优势;另一方面,泛在网络与通信技术、传感器技术、RFID技术、云计算技术的发展,使得未来信息化呈现出新的发展方向和模式,使旅游信息的精确采集、旅游信息资源化应用成为可能。

2.3 智慧城市技术体系架构

智慧城市的应用计算呈现分布式的特征,各智慧行业的专业应用处理由行业私有云计算中心完成。在总体设计部分,我们抽象了智慧城市技术体系架构的主要4个层次:数据汇聚层、数据活化层、行业应用层及应用服务层。

2.3.1 总体技术架构

智慧城市的技术架构必须能够促进跨政府部门的信息共享及跨政府部门应用协同,并能够通过信息服务实现城市级综合信息查询,通过应用服务支撑智慧应用的快速开发与运行,在这个技术平台上为各类智慧应用提供统一的运行管理,并通过先进的云计算技术,促进城市级基础设施共享服务和平台服务,如图2-4所示。

功能平台主要由信息平台、城市信息库及城市云计算中心这3个部分组成:

(1) 信息平台:信息平台主要提供信息的共享、交换、查询等方面的服务,并提供相应的信息数据接口以支撑智慧城市的应用,需要提供一个城市级别的跨行业信息平台 and 各个不同行业的相应信息平台。它致力于提供智慧城市可公开的各类数据资源及服务,为企业、科研院所和社会公众开展政府信息资源的社会化开发利用提供数据支撑,推动智慧城市行业应用的开发及服务。

(2) 城市信息库:它是智慧城市信息的集散中心,疏导着各智慧行业应用的需求,按请求的领域、行业的层次、粒度来进行数据路由。城市信息库将由各智慧行业应用库的索引元数据库、城市基础数据库和行业主题库三大信息库构成。其中城市索引元数据库是由各智慧行业应用库抽取、清洗、汇总所得,是产生城市基础库和行业主题库的前提。城市基础库包含人口库、法人库、地理信息库、资源库。城市行业主题库是对索引元数据库和城市基础库按目标主题纬度的切割立方体,为跨部门、跨行业的特殊目标主题应用服务。智慧城市的信息按数据生命周期原则进行统筹管理,按数据性质分级管理,按数据类型规划存储方式,按数据粒度规划存储介质。

(3) 城市云计算中心：城市云计算中心广义上可认为是一种云计算信息服务平台，狭义上也可理解为是一种城市公有云数据中心。它将不再如传统数据中心般服务于单一行业应用，它更像是智慧城市信息高速公路的交通枢纽，串联中转各智慧行业私有云应用；它又像是一种目录服务，以汇总元数据模式统筹管理着各智慧行业应用数据，它更像是城市的智慧大脑，通过类似神经组织般的泛在城市网络连接智慧城市的每个业务单元，并以信息为载体，感知、控制、管理、决策着智慧城市的方方面面。城市云计算中心是智慧城市泛在网络的交换中心，是智慧信息的集散中心，是应用计算的调度中心，是信息安全的预警中心。



图 2-4 智慧城市总体技术架构

智慧城市技术平台需要提供关键的 6 大核心能力：

- ① 快速应用服务提供能力：快速应用服务的提供依赖成熟的支撑平台和模块化组件。
- ② 城市泛在网络覆盖能力：城市网络互联采用横向整合和纵向延展相结合的规划思路。横向整合即是全面融合已经建成的政务专网、无线网、平安城市网、行业专网等内容。纵向延伸物联网、传感网，依托互联网延伸至企业、社区、家庭。

③ 第三方系统的集成能力：随着物联网技术在智慧城市中的深入运用，专业外设、外部专用系统需要集成，智慧城市的一体化特征要求串联各智慧行业。智慧城市中央平台应是一个开放的、标准化、自适应的平台，与外部系统在通信接口、数据结构、控制指令环节无缝融合，通过消息传递的服务总线串联。

④ 数据统一管理分析能力：分布与集中并存，元数据和基础数据的统一，主题数据切片，数据的循环生命周期管理等。

⑤ 城市系统资源共享能力：基础设施服务的平台化，按需的分配和回收机制，计算，存储，网络资源的复用，安全服务的协同。

⑥ 城市系统平滑演进能力：城市云中心是城市信息化的演进而非颠覆，城市云中心并非取代各行业现有系统，而是促进跨行业、跨领域的数据共享与业务协同。各智慧行业汇聚业务应用的基础元数据于城市中央数据库，形成城市级公共应用，弥补跨行业应用空白，促进城市主题应用发展。

2.3.2 开放性体系

1) 开放性设计

智慧城市的技术体系架构的开放性设计中，根据物联网的体系架构提出智慧城市发展要点：前瞻考虑、整体设计、自上而下、全局应用和架构的规划，面向应用聚合演进考虑。我们将所有政府部门和产业单位包容其中，各部门必须进行产业协同。整体开放式架构智慧城市方案的设计可以从3个平面进行概括(图2-5)：

(1) 第一个平面是整体体系架构设计。整个网络中的智能处理、计算能力分布在3个层面上，包括感知层的感知层面、数据层的数据活化层面以及应用层的应用基础平台层面。通过3层的智能分布与相互关联，实现整个系统的能力均衡部署，构成一个明确的、标准的、面向未来可自由扩展的开放式体系架构。并在此架构基础上提供一个弹性的业务环境——可以自由地添加、删除业务，业务之间可以联动和协同。

(2) 第二个平面是整体应用的布局。包括智慧城市管理领域、智慧市民服务、智慧的宜居设施等，此外需要更多地考虑城市产业的发展。在具体的行业方面，涉及智能楼宇、智能家居、路网监控、智能医院、城市生命线管理、食品药品管理、票证管理、家庭护理、个人健康与数字生活。因此智慧城市需要在具体的城市里实现示范应用，比如在海南可以提出体验式的旅游；在深圳可以提出打造一个高科技园区、品牌产业的组织等一系列模式。通过不同领域的相互关联和交叉，来实现应用的整体布局。

(3) 第三个平面是运营管理的设计。通过统一的平台，可以在业务层面解决业务和资源运营管理的问题，在网络层面解决网络抽象和资源管理的问题以及应用的策略和控制等。在终端层面，除了终端和网关的管理，还会把管理的手段延伸到感知层，通过对一些传感器的管理，真正实现整个体系架构从

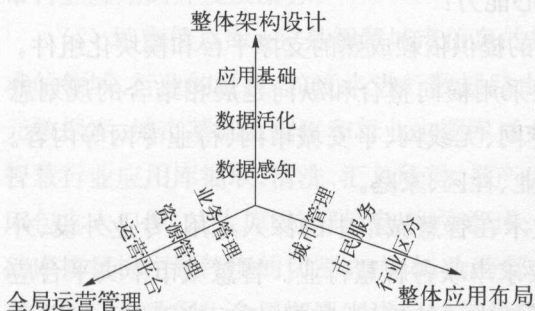


图 2-5 智慧城市技术体系架构开放性设计原理图

业务到网络、到终端、到感知的完整的端到端的可运营、可管理的系统。

智慧城市的技术体系架构是一种实用的、有条理的部件集合的结构化形式,该架构通过这些部件以一种精确的方式为政府、企业和个人用户提供远景的支持。技术体系架构的实现连接了具体的需求和远景的战略规划。从本质上讲,技术体系架构定义了智慧城市为了获得长期的应用前景而构建与使用的信息技术平台。

技术体系架构并不仅仅是一套标准,确切地讲,它是一种组织能力的体现,这种能力不仅体现在文档方面,还反映了技术专家、IT 管理员、规划设计人员以及实施人员的知识与经验。智慧城市的技术体系架构的开放性依赖于标准的开放性和接口的开放性。

2) 可扩展性设计

可扩展性(Extensibility)是智慧城市技术体系架构设计的另外一个难点,可扩展性要求我们提出的智能城市技术体系具备灵活而便捷的扩展方式。在通常情况下,可扩展性必须满足如下的 3 种特性:

(1) 松耦合性:为了控制智慧城市技术体系的复杂性,因此整个系统必须在逻辑上进行划分,各个逻辑块之间应该是松散耦合,需要对不同领域的理解和仔细分析。

(2) 可重构性:智慧城市技术体系架构随着时间的变化而发生变化,因此可扩展性不是一蹴而就的,需要随着对相关业务领域理解的深入而不断重构获得。可重构计算理论上是指在任何信息处理系统中,软件和硬件模块都可以根据变化的数据流或算法进行重新配置或重新设置。重构可以发生在设计阶段、运用阶段、两个执行阶段之间或执行过程中。这些时间段的每一个都定义了一种独特的可重构系统类别。

(3) 可演化性:在智能城市的技术体系架构中,可以预知这个系统会根据应用需求和网络环境的变化而发生动态演化,这些演化体现在构成其实体元素数目的可变性、结构关系的可调节性和结构形态的动态可配置性。因此,可扩展性应该具备适应开放、动态和多变环境的应变能力。在系统的演化过程中,时刻准备着对复杂性保持关注,确保这些复杂性得到消化。

智慧城市的技术体系架构的可扩展性设计中,智慧城市是新一轮信息技术变革和知识经济进一步发展的产物,是以互联网、物联网、电信网、无线宽带网等网络的多样化组合为基础,更加广泛而深入地推进基础性与应用型信息系统开发建设和各类信息资源开发利用,形成的技术集成、综合应用、高端发展的网络化、信息化、智能化和现代化城市,是以智慧技术、智慧产业、智慧人文、智慧服务、智慧管理、智慧生活等为重要内容的城市发展新模式。

关于智慧城市技术体系架构的可扩展性,应该考虑到智慧城市研究能够自我演进、适应智慧城市建设的开放式、可扩展的体系架构和逻辑架构,满足城市基础差异化、城市行业多样化、城市信息多元化和城市管理智能化的需要。同时,考虑到智慧城市的体系构架响应评测和具体服务结果的反馈机制,根据反馈信息自我演进,形成一个良性循环的有机整体。

还需要考虑智慧城市开放式应用接口技术,做到对不同应用领域、不同技术的高度兼容。重点研究体系架构对应用接口的良好支撑、应用接口与体系架构其他部分的整合;研究智慧城市体系架构与现有技术体系的无缝衔接,实现各种技术在智慧城市体系架构下的高度可移植性,降低各种开发成本。

智慧城市技术体系架构可扩展性的设计可以从3个方面进行概括:

(1) 第一个平面是在技术体系架构的模块层次上,可扩展性要求满足松耦合性。因此,整个5层体系结构中,层与层之间使用松耦合结构,每一层对上层提供的服务采用透明机制,用户只要明确下层服务的调用接口,就能够容易地访问到相应的服务;同样的道理,在每一层次上,技术模块之间也采用松耦合的方式,彼此之间的依赖性要降低。在此基础上,对于新增加的技术模块的设计也采用相同的思想,提供一个弹性的业务环境。

(2) 第二个平面是在技术体系架构的接口层次上,可扩展性要求满足可重构性。设计难点在于修改已发布的接口,因为已发布的接口会供外部人员使用。因此,修改接口会导致引用接口的其他技术模块不修改程序就无法运行。修改接口最好的办法是增加一个新的接口,让旧接口调用新接口,这样原来的程序就不用修改了。此外,还要考虑因为修改接口带来的新问题,往往对于接口的另一个建议是尽量不要发布接口。另外一点,需要实现动态重构性。因为在智慧城市的整体技术架构中,涉及不同行业、不同应用,因此在突发事件的瞬时演化、衍生及生命周期过程中,表现出应急资源需求的不确定变化,导致应急联动组织的动态重组响应。

(3) 第三个平面是在技术体系架构的应用业务层次上,可扩展性要求满足可演化性。解决智慧城市技术体系架构的可演化性,除了要解决模块设计和实现的正确性外,还必须解决体系架构周期长、复用性差、成本高、文档多以及难以适应系统演化等问题。可以采用面向服务(SOA)的架构来实现可演化性,它是使用标准化的技术,与特定的技术平台无关,而几乎所有的技术平台都支持这一技术标准。

2.3.3 城市互联平台架构

智慧城市的技术架构需要考虑到彼此之间的互联互通,这样才能做到区域经济的协调同步发展,在设计智慧城市技术体系的时候,一定要考虑不同智慧城市平台之间的协作与互联的机制保障。在这里主要考虑基础数据注册、目录服务、实时检索及安全策略4个部分的内容,如图2-6所示。

1) 基础数据注册

由于接入云互联平台的应用多种多样,而且对于重要标识(如公民、公司)的表述也各不相同,需要在云互联平台中提供对独立实体标识的统一性表述。这需要包含以下各个功能点:

(1) 实体数据的清洗及匹配,在互联平台中存在并只需存在一个实体表述;这里有一个

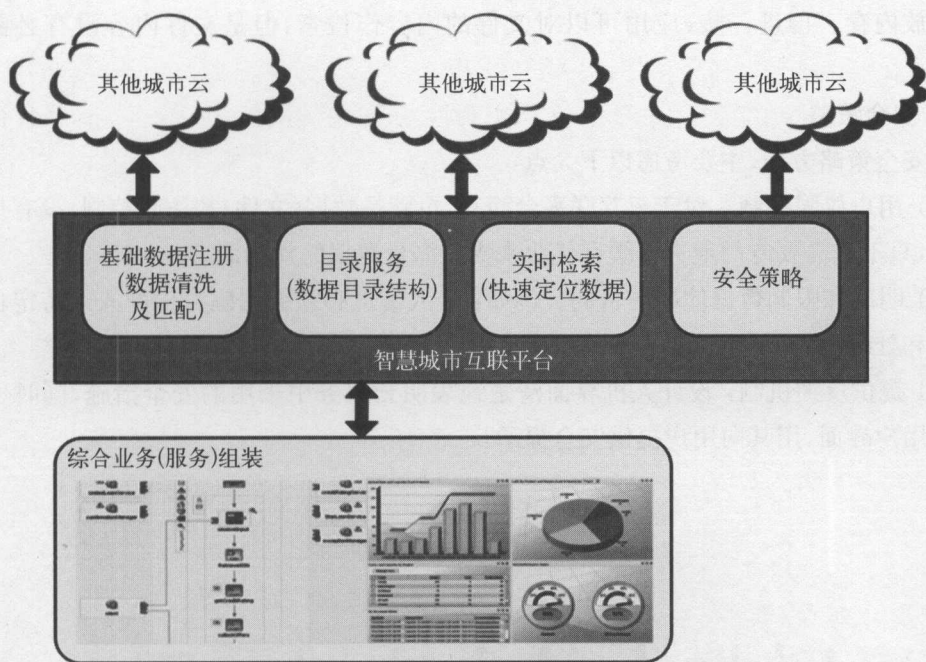


图 2-6 智慧城市互联平台

原则,发送和存储尽量少的业务信息到云中。通过对系统的分析,只对最小一部分业务信息进行必要的收集和储存,这样做的好处是能使数据在存储和处理过程中只需采用必要强度的保护措施;所处理的数据可以采用相关的加密技术来进行保护。

(2) 对于实体表述的各个部分的维护更新由特定的行业应用进行维护(例如:某个公民的犯罪记录只能通过公安系统进行维护,其他系统只有阅读权限)。

(3) 实体基础数据的表述需要具有一定的弹性,可以考虑使用弹性域(Flexible Field)特征构建。

2) 目录管理系统

海量数据的目录管理是一个困难的课题。目录管理主要应用于组织用户处理海量数据时,理论上,不会在单个目录下放太多的文件(如果数据太多,在展现时考虑使用分页技术)。这里可以采用非关系型数据库(NoSQL)存储方案来达成需求。它对存储的数据,会提供基于字典库的游标查询。这需要包含以下各个功能点:

- (1) 游标的遍历及快速定位准确信息。
- (2) 字典库的自动学习(指智能的维护目录字典库的算法)。
- (3) 基础目录建立的功能。

3) 实时检索

实时检索是建立在海量数据平台上的一个重要的搜索应用,需要有一个传统的检索系统,然后想办法提高检索的实时性。可以设计一个内存索引,把用户新增的文件,对文件名切词后放到内存中检索,检索的结果参与最终的合并。每隔 5 min 合并到传统检索系统中,

然后释放内存。再进一步,应用可以对文件的内容作检索,但是文件内容没有必要做到实时。

4) 安全策略

在安全策略方面,主要考虑以下3点:

(1) 用户权限控制。对于云互联平台而言,也就是数据(文件)的访问控制,没有什么特别之处,只不过需要专门地进行缓存优化来支持海量数据的判别请求。

(2) 明确和限制数据使用的目的。例如:个人信息必须被明确身份的人以特定目的进行使用和处理,而且使用前要联系信息拥有者。

(3) 提供反馈机制:设计人机界面清楚地表明云服务中采用的安全措施,同时可以设计图形用户界面,用其向用户提供安全提示。

◇参◇考◇文◇献◇

- [1] 中国电子技术标准化研究. 中国智慧城市标准化白皮书. 2013.

第3章

智慧城市中的大数据资源

智慧城市中的大数据资源包括：智能交通、智能医疗、智能楼宇、智能电网、智能农业、智能安防、智能环保、智慧旅游、智慧教育、智能水务等大数据资源，涉及众多智慧城市应用范畴。

对于大数据，企业和学术界目前尚未形成公认的准确定义。维基百科将大数据定义为“无法在一定时间内用常规软件工具对其内容进行抓取、管理和处理的数据集合”；权威 IT 研究与顾问咨询公司 Gartner 将大数据定义为“在一个或多个维度上超出传统信息技术的处理能力的极端信息管理和处理问题”；美国国家科学基金会(NSF)则将大数据定义为“由科学仪器、传感设备、互联网交易、电子邮件、音视频软件、网络点击流等多种数据源生成的大规模、多元化、复杂、长期的分布式数据集”。尽管存在不同的表述，但一个普遍的观点是，大数据与“海量数据”和“大规模数据”的概念一脉相承，但其在数据体量、数据复杂性和产生速度三个方面均大大超出了传统的数据形态，也超出了现有技术手段的处理能力^[1]。

大数据包含了互联网、传感设备、视频监控、移动设备、智能设备、非传统 IT 设备等渠道产生的海量结构化或非结构化数据，并且时时刻刻都在源源不断地渗入城市日常管理和运作的方方面面。

大数据的特点为体量巨大、有价值信息的密度低，因此，需要对海量数据进行分类整理，形成一定的数据资源库。数据资源库是通过大数据分类技术，按照用户预设的类别体系，将数据进行归类所形成的资源库。由于实际的数据处理业务往往面临着不同的分类需求，因此数据资源库的建设需要采用不同的分类依据，例如：数据量大小、业务性质等。在实际业务中，数据资源库的分类依据是多种多样的。

(1) 按照数据资源的公开程度分类，数据资源库分为可公开的、在一定范围内可公开的、私密的(非公开的)三类；此处加入国家对数据库安全方面的规定，强调在实际管理和使用过程中需要遵循相关法规对数据安全的约束。

(2) 按照数据资源的用途分类，数据资源库分为提供公共服务的、特定部门使用的(管理、监督、决策)两类。

(3) 按照数据资源的性质分类，数据资源库分为原始录入的数据、经过一定处理分析的数据库两类。

(4) 按照数据资源使用单位的性质分类，数据资源库分为商用(民用)、政府使用两类。

本章主要介绍智慧城市数据资源体系架构与智慧城市数据资源交换体系和目录体系。其中，智慧城市数据资源体系架构主要介绍数据资源体系、数据资源存储、基础数据资源库、共享数据资源库、特定数据资源库、公共服务数据资源库、管理数据资源库和政府决策数据资源库；智慧城市数据资源交换体系和目录体系主要介绍数据资源交换体系和数据资

源目录体系。

3.1 智慧城市数据资源体系架构

智慧城市融合了最新的云计算、物联网、互联网、移动通信等新一代信息技术,通过智慧的感知、分析、集成为政府在行使经济调节、市场监管、社会管理和公共服务政府职能的过程中提供一个优化的基础设施、融合的信息资源、协同的业务应用环境^[2]。涉及环境监测、城市交通、公共服务、居家生活、经济商务、健康管理、公共安全等诸多领域,支撑这些不同领域智慧应用的是来源广泛的各类数据,这些数据资源是保障智慧城市建设的

基础。

一方面,智慧城市管理需要深度整合、分析数据。政务信息化建设的推进,在政府各行业、部门里建设了一大批不同专业领域的应用系统,这种应用系统解决了政府部门内部专业领域的应用需求,并产生了大量专业的数据。同时,这种局面造就了很多信息孤岛,而这种孤岛之间是相互独立的,部门之间数据共享相对困难。信息共享困难所造成的政府决策缺乏数据支持、数据分析力度不够充分所带来的问题亟待提高。信息资源作为一种资源,在经济社会发展中应该被充分利用起来,为政府管理部门决策提供信息依据,并提供重要的辅助决策依据。通过建立统一的元数据标准规范梳理元数据,进行数据的整合,形成数据资源,并加以分析与利用。在城市运行管理中,为政府管理决策提供重要的数据依据,也是提升政务管理水平的重要手段^[3]。

另一方面,大数据环境下的智慧城市运行体征管理,需要对城市运行状况进行信息的收集和整合。通过基于整合的数据信息对城市各领域的运行状况进行监测,并且跟踪分析城市的运行体征状况进行管理,实现城市管理工作的改进和优化。基于数据的智慧城市运行体征管理是进行城市管理一种理念,通过基于数据信息的整合、数据的展现监测分析、发现问题、制定措施与执行、继续监测分析等形成一个可持续、健康发展的闭环过程,为智慧城市运行管理的可持续发展与改进提供数据支持,例如^[3]:

(1) 在城市管理中,保障城市基础设施的正常运转是城市管理运行保障工作的首要任务,通过移动巡检等物联网应用对城市运行中的基础设施实时监测数据与业务管理数据进行整合,并按照业务目标构建指标体系,梳理设施运行体征,疏通不同网络之间的数据流动,为城市设施体征提供数据监测,为设施应用分析提供支撑组件和管理功能,充分发挥数据潜在价值,从而提高政府的城市基础设施监管水平。

(2) 政府管理层知悉社会公众的民情、民意对于政府管理和掌握百姓真实意愿至关重要,通过掌握民情、民意,了解老百姓的愿望和诉求,才能更好地进行城市管理。采用民情热线、舆情监测、移动终端等物联网技术对城市中的人、事、情、组织产生的民情、民意数据

进行整合。通过构建民情指标体系,梳理民情运行体征,并加以分析,充分发挥民情数据的潜在价值,为政府管理提供数据参考,提升对公众的服务水平。

(3) 公共安全从来没有像今天这样遭受到残酷的考验和严峻的挑战,人们对公共安全的现在和未来充满顾虑。对城市运行中的公共安全相关数据进行整合,并构建公共安全指标体系去量化和考评公共安全的管理工作,通过梳理城市安全运行体征,进行基于整合数据的安全监测、综合分析、预测预警,为城市安全运行提供数据服务,提高城市公共安全的管理水平。

目前,随着各行各业信息化建设的开展和信息化水平的提高,各类信息资源日益丰富。但是各信息化项目所产生的信息资源由于在建设过程中缺乏统筹规划,缺少统一的数据标准和规范,存储分散,缺乏共享或共享服务质量与效率不高,并且缺乏深度开发和综合利用,相关系统建设的时间成本、经济成本较高。城市建立在分割的领域之上,城市管理者所需的全局关键信息往往被埋藏在各个城市机构独立部门的不同系统中,他们无法获得管理所需的清晰视图。城市日常管理面临错综复杂的业务流程以及各种盘根错节的业务系统,但处理城市级的问题往往需要大量历史和实时的相关数据,形成 360°的全景视图,这才有助于城市管理者作出明智的决策。因此,智慧城市有关大数据的应用和分析存在很多问题^[4]:

(1) 城市管理者无法获知实时状态,并进行实时地整理和分析,同时要能够下达相关的指令对各相关单位进行调动和指挥。

(2) 既有系统针对专门领域,不关注其他领域应用。各机构的工作,没有对事件、事故或潜在危机的整合性单一视图,就无法快速分享信息,在事件规划与执行中,缺乏领域间的协作与交流,很难整体协调。

(3) 城市每天的运作产生了大量来自不同渠道的数据,但城市常常缺乏可视化且无法抽取有用信息。

因此,需要从整合现有信息资源并为未来的信息化建设提供统一支撑的角度,打破各个系统间数据不能共享的现状,实现空间信息和非空间信息的统筹管理、统一服务、共享共用。针对共享的信息资源,通过专业的基础架构实现不同数据的快速查询、统计、关联等操作,为政府各级领导和各业务应用对象,提供有效支撑业务应用、领导决策和内外部的信息共享^[5]。这就要求有针对性对数据资源进行接入、存储、处理、融合、应用的智慧城市数据资源处理平台,以统一标准和数据安全措施提供数据融合应用服务,其架构如图 3-1 所示:

智慧城市的数据来源广泛、结构多样,除了已有的各业务系统的数据以外,还有来自物联网、互联网等的实时、非结构化数据,因此需要不同的数据交换方式以接入这些异构数据。根据数据的特点存储在不同的数据库,根据应用服务的需要进行数据融合处理工作,从而为上层服务提供数据基础。

因此,数据资源处理平台是智慧城市建设的灵魂,它搭建了对外提供应用服务、对内实现整合的一体化数据采集、数据资源管理和信息服务的业务服务平台,将不同来源、不同格式、不同类型、不同应用的数据进行规范、融合,形成智慧城市的数据资源体系,并对外提供

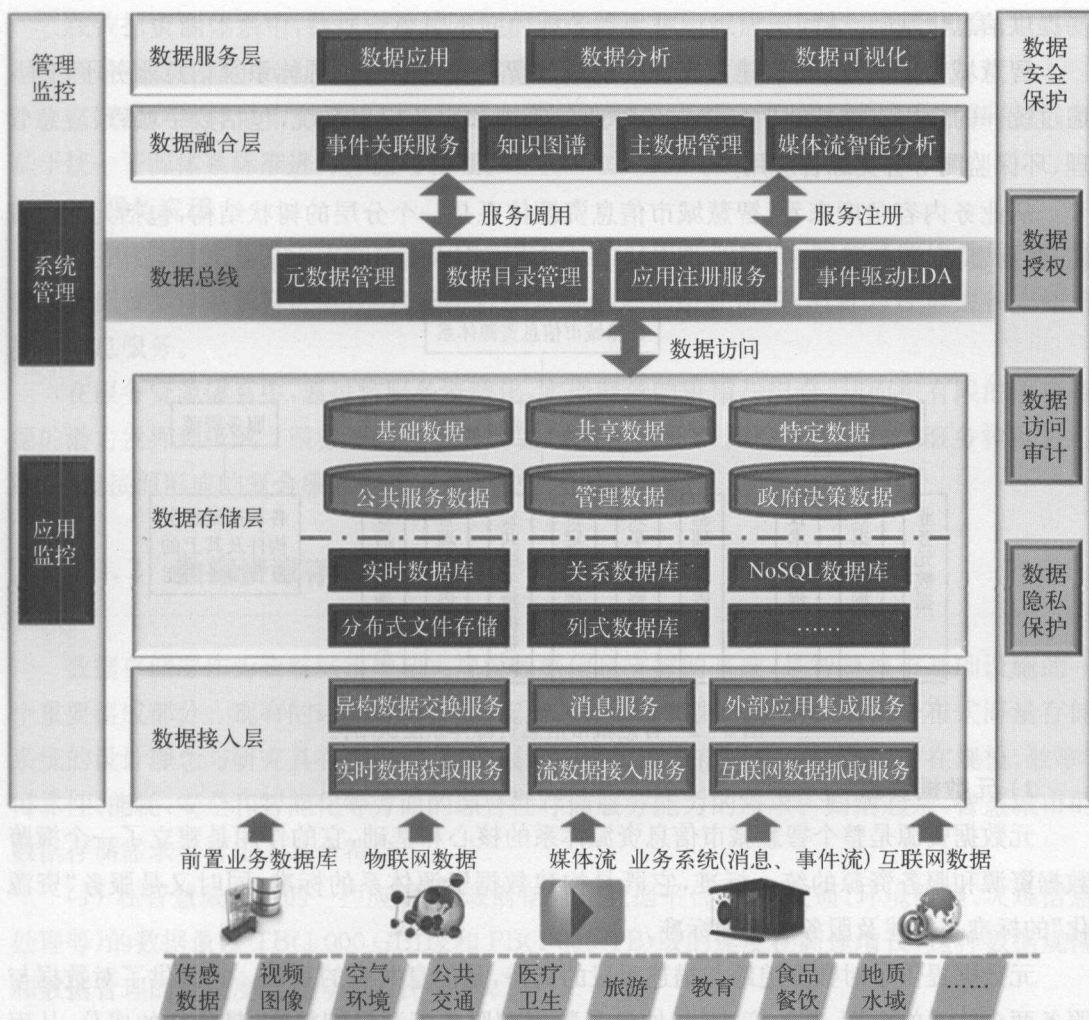


图 3-1 智慧城市数据资源处理平台

统一的数据共享和信息服务^[6]。

3.1.1 数据资源体系

数据资源体系是智慧城市的统一公共信息共享平台。这个平台有两层含义，一是对分散在各个专业业务系统中的原始信息资源进行集中整合，形成统一的信息资源体系；二是对各种信息资源进行适当封装，以组件化的模式，向包括决策层领导、各部门管理人员和广大社会公众在内的不同层次用户，提供内容丰富、形式多样的综合信息服务。

有了这样一个统一公共信息共享平台，就能够把所有的城市业务管理与服务系统纳入一个统一的信息系统框架，真正实现“一盘棋”效应，彻底改变传统管理体制下条块分割、各自为政的局面，打破信息孤岛，为建立新的开放式、交互式、协同式城市管理 with 民生服务格

局提供信息化基础支持。

智慧城市的实现,需要建立在标准规范的智慧城市信息资源体系及信息服务平台上,通过统一的应用支撑基础平台建立各种业务管理和应用系统,包括数字城管、应急管理、环保监测等各类综合应用^[1]。

从业务内容角度来看,智慧城市信息资源体系是一个分层的树状结构,包括元数据资源、数据资源、服务资源三大类。具体业务内容层级结构如图3-2所示。

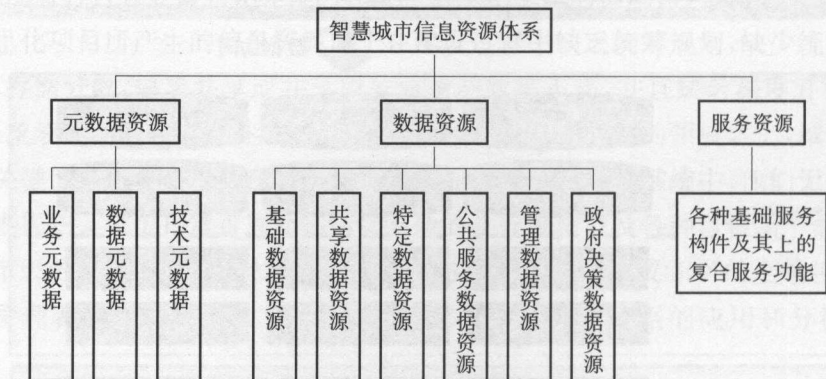


图3-2 智慧城市信息资源体系层次结构

1) 元数据资源

元数据资源是整个智慧城市信息资源体系的核心和基础,它的作用是建立了一个横跨数据资源和服务资源的统一标准,它既是构建数据资源体系的标准,同时又是服务“资源化”的标准与描述及服务的管理标准。

元数据提供了对数据追踪与描述手段的支持,是信息共享的基础。它提供了对数据与服务两个体系的全面、细粒度描述,抽象了数据和服务资源最细粒度、最稳定的成分,从而为管理“精细化”提供了基础支持。

元数据资源中,包含三类描述不同层面内容的元数据:

- (1) 业务元数据:与用户业务紧密联系的反映业务实体对象本质业务含义的元数据。
- (2) 技术元数据:是与技术实现相关的反映业务实体对象技术属性的元数据。
- (3) 数据元数据:描述与业务实体对象的数据信息相关的业务特征与技术特征的元数据,偏重于描述数据本身的特点。

元数据涉及系统的所有资源和所有业务阶段,对系统及组成系统的各种资源的管理,很大程度上是通过对元数据的管理来实现的。

2) 数据资源

数据资源体系是智慧城市信息资源体系中各种数据资源的总集合,集数据的“累积”和“管理”于一身。数据资源体系不断地“汇聚”各式各样的数据,并把它们转化为统一标准的数据“累积”下来,并统一以资源的形式进行信息的管理与组织。

在数据资源体系中,数据资源按不同的服务需要被组织成不同粒度的数据集,这些数据集既是资源管理单元也是资源服务的信息提供单元。

在数据资源体系中,无论是外部获取的原始数据还是在自身体系内产生的新信息,都基于统一资源体系标准进行存储与管理。

3) 服务资源

服务资源体系是基于数据资源为用户提供的信息服务的集合,服务资源体系的作用是展现数据资源体系现有价值,深入挖掘数据资源体系的潜在价值,有针对性地为各类用户提供信息服务。

在服务资源体系中,通过将服务资源化,实现服务的重用与组合,从而把有限的服务资源的潜力发挥到极致。服务资源包括各种基础服务构件以及基于这些基础服务构件进行组合、连接而形成的复合服务构件以及各种应用服务功能^[1]。

3.1.2 数据资源存储

数据存储层作为智慧城市中超大规模数据信息支撑的基础,是智慧城市基础设施的一个重要组成部分。现有的以大规模、高性能与可扩展为主要基础的网络化分布式海量存储系统的设计理念与研究具有局限性,已经很难满足智慧城市中多种复杂应用在规模、效率、可靠性、能耗、安全和智能化等方面的综合性存储服务能力的需求。归纳起来,智慧城市的数据存储需求具有以下几个特点:

(1) 在智慧城市中的一些应用(如政府信息化数据平台、数字交通、环境监测、灾难信息处理等)的数据量以 TB(1 000 GB)级和 PB(1 000 TB)级的速度在高速增长,这种数据规模和数据管理的复杂度已经今非昔比。

(2) 各种应用对数据的可访问时间和访问持续时间在增加,使得数据必须保证能够时刻处于可访问状态。

(3) 数据种类和数据应用的不断繁衍,海量数据呈现动态生长、异构多样化、地理分散、语义丰富等特征。

(4) 数据应用的用户分布广泛,用户访问方式存在多样化与群体众多性。

(5) 应用需要可靠、高效低耗与经济性相结合的数据存储。

针对智慧城市中的数据资源存储需要考虑到:

(1) 分布式大数据架构。智慧城市在日常应用中会产生视频、图片、日志信息、文本信息等大量数据,并且随着时间的推移,数据量增长速度极快。通过对这些海量数据的分析和处理,提取出有价值的信息,是智慧城市的政府决策、工业经济发展、公共安全、城市应急防控、社会公共服务等方面发挥作用的前提。因此,需要采用分布式存储方式,提高读写速度,扩大存储容量。

(2) 大数据存储与管理。大数据给传统的数据管理方式带来了严峻的挑战,关系型数

数据库在容量、性能、成本等多方面都难以满足大数据管理的需求。云数据库可以不受任何部署环境的优势,随意地进行拓展,进而为客户提供适宜其需求的虚拟容量,并实现自助式资源调配和自助式使用计量^[7]。

(3) 高可扩展性(图 3-3)。智慧城市中数据存储的按需视图特性需要高可扩展性的支持,不仅要为存储本身提供可扩展性(功能扩展),而且必须为存储带宽提供可扩展性(负载扩展)。云存储的另一个关键特性是数据的地理分布(地理可扩展性),支持经由一组云存储数据中心(通过迁移)使数据最接近于用户^[8]。对于只读数据,也可以采用多副本和数据布局技术来提高访问性能。

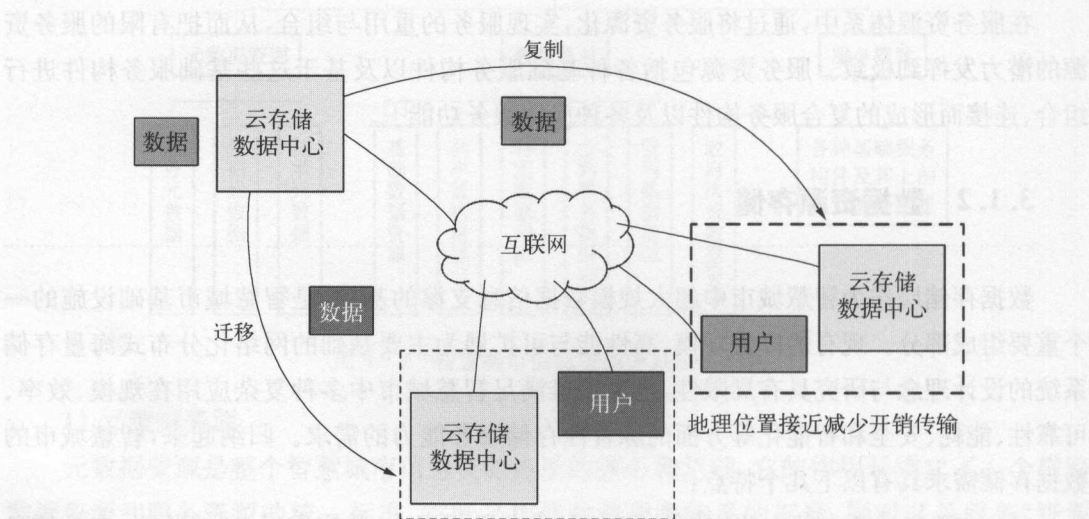


图 3-3 可扩展的弹性云存储

3.1.3 基础数据资源库

基础数据资源库的概念是 20 世纪 90 年代电子政务产生后出现的。电子政务是指运用计算机、网络和通信等现代信息技术手段,实现政府组织结构和工作流程的优化重组,超越时间、空间和部门分隔的限制,建成一个精简、高效、廉洁、公平的政府运作模式,以便全方位地向社会提供优质、规范、透明、符合国际水准的管理与服务。智慧城市依托基础数据库系统的建设,可以实现我国政府各委办局、各所辖地区的经济社会综合数据采集交换,为各部门提供更广泛的信息共享支持,一方面数据信息从各委办局、各所辖地区整合接入,另一方面也为政府和这些接入部门提供全面的共享服务。同时,以智慧城市基础数据库指标体系建立为基础,整合来自各委办局和各所辖地区的、经过审核转换处理的数据资源,可实现对经济社会信息的统一和集中存储,确保数据的唯一性和准确性,为今后政府工作提供一致的基础数据支持。

基础数据资源库有以下特征:首先,它是公开的;其次,数据库设立的目的是为了提供

公共服务;第三,数据是原始录入的;第四,数据资源库的使用者是全体社会公民。

2002年,我国国务院信息化领导小组第二次会议发布了17号文件《国务院信息化领导小组关于我国电子政务建设指导意见》,该文件将人口、法人单位、自然资源和地理空间、宏观经济基础数据库这4个数据资源库作为国家基础信息数据库,列入国家电子政务建设的重点任务。

1) 人口数据库

该库基于城市级的政务网络、空间地理信息系统以及支持多业务部门的数据集成和交换平台,按照统一的数据库编码标准,收集、比对、整合分散在各部门的基础数据,集聚、整合各参建部门管理系统中涉及人口的基础信息,建立人口基础信息库、集中与分布相结合的人口数据库体系,为人口信息的共享交换提供基础。

人口数据库人口信息包括本地常住居民身份信息、流动人口信息、已婚育龄妇女信息、社保缴纳和发放信息、居民收入基本信息和享受低保人员信息等人口生命周期各个阶段产生的所有信息。通过建立这样一个全面的人口数据库,可以方便公安、计生、民政、卫生、教育、残联、人力资源和社会保障、统计等使用人口信息的部门及时准确地查询到人口相关信息,不同部门可以共享相关的人口信息,确保各相关部门人口基本信息的准确性和一致性。实现人口计生部门获取其他部门提供的新婚、出生、死亡、劳务、流动和总人口等信息,对系统漏统信息和不准确信息查漏补缺,提高人口数据库的完整性和准确性,可以更好地为群众提供优质服务和人性的管理;教育部门掌握其他部门统计的各年龄组人口数据,对未来教育结构的调整及早谋划;公安部门可以获取其他部门提供的出生、死亡、流动人口和总人口信息,补充和完善本系统数据库,提高出生人口的上户率和死亡人口的销户率;民政部门可以预测未来新婚对象的人数变动趋势;人力资源和社会保障部门能够掌握其他部门提供的流动人口数量流向,合理引导劳务输出;卫生部门获取其他部门提供的出生和死亡信息,提高住院分娩率和数据质量;统计部门获取其他部门提供的出生、死亡和流动人口数据,提高抽样调查数据的科学性和准确性。同时人口数据库及交换平台应满足为市民卡项目、数字医疗、数字校园、数字社区等项目提供人口数据支撑。

2) 法人单位数据库

该库基于城市级的政务网络、空间地理信息系统以及支持多业务部门的数据集成和交换平台,按照统一的数据库编码标准,收集、比对、整合分散在各部门的基础数据,集聚、整合各参建部门管理系统中涉及法人的基础信息(如机构代码、机构名称、机构类型、经济行业、业务经营范围、机构地址、法定代表人等字段以及机构变更、注销的相关信息),建成标识统一、结构科学、查询快捷、动态管理的法人基础信息数据库;制定与交换平台对应的相关标准、制度和规范的安全运行管理体系;实现工商局、地税局、国税局、质量技术监督局等法人数据相关业务部门之间的网络互联和业务数据的实时交换与应用。为政府建立“信息共享、业务联动、交叉稽核、统一监管”的经济管理和企业服务体系提供技术手段,为行政审批系统等项目提供法人单位数据支撑。

3) 自然资源和地理空间数据库

该库包括空间信息展示子系统和信息服务子系统。前者主要用于向各部门发布资源和地理空间信息,使得各部门工作人员可以通过 Web 形式对数据库中的信息资源进行浏览检索,了解可共享的信息资源;后者则主要是由一系列数据服务接口组成,提供的数据共享服务主要有目录服务、电子地图服务、相关信息图层服务、遥感影像服务、图片多媒体数据服务等。

自然资源和地理空间数据库的建设要满足以下要求:尽量减少空间数据存储冗余;提供稳定的空间数据结构,在用户的需要改变时,数据结构能够作出相应的变化;满足用户对空间数据及时访问的需求,高效提供用户所需的空間数据查询结果;在空间元素间维持复杂的联系,反映空间数据的复杂性;支持多种决策需要,具有较强的应用适应性。

4) 宏观经济基础数据库

该库主要涉及宏观数据、行业数据、企业档案、企业名录以及其他用户所需要的信息服务。其中,宏观经济信息库囊括了各种类型的宏观经济数据指标,帮助用户准确了解国家宏观经济动态,把握经济运行大局;行业基本数据库囊括了主要行业产品产量、进出口和产销率等数据信息,帮助用户掌握行业最新产品动态,作出准确的生产规划和战略决策;行业经济效益库囊括了主要行业经济效益指标,帮助用户掌握行业整体运行情况、盈利状况,做出正确的发展规划;企业绩效评价库给出了各主要行业企业绩效水平的参考值(优秀值、良好值、平均值、较低值、较差值),是权威、全面地衡量企业管理运营水平的评价标准。此外还包括重要的企业档案。例如:企业注册资料、主要管理者及其背景、股权结构、主要经营业务、资产负债表、损益表等信息。

宏观经济基础数据库有助于用户及时准确掌握宏观、行业经济运行态势,研究制定年度计划和中长期发展规划,准确掌握竞争对手、合作伙伴、目标客户的业务状况,主动调整合作方式,降低经营风险等。

值得一提的是,上述4个国家基础信息数据库并不能涵盖智慧城市基础数据资源库的全部。智慧城市的建设涉及区域内经济社会发展各个方面的数据,其基础数据资源库的建设需要从所有相关政府部门采集数据。选择和确定智慧城市基础数据资源库项目共建单位的主要依据包括:

- (1) 共建单位拥有的基础指标在反映地区经济社会发展进程和地区城市指标体系中的重要程度。按照设置指标的依据和原则,参与共建单位的基础指标必须能够较全面地满足要求。
- (2) 各方面对指标的需求程度,包括政府领导对相关信息的关注程度,有关部门的需求程度等。

3.1.4 共享数据资源库

共享数据资源库的显著特征是其数据资源能够被至少一个以上的部门所共同利用。

资源是从经济学视角提出的概念,能够被人类有效利用的物品才有价值,才能谈到稀缺,才需要珍惜,资源的概念才会产生^[9]。数据资源库服务被视为信息服务,它按照用户的检索要求直接提供数据处理结果,不需要用户准备数据也不需要用户处理数据。数据资源库查询依据用户目标对数据资料进行选择加工,是一种信息的挖掘、提升信息浓度的服务。数据资源库服务面向最终用户,强调服务结果,强调对用户的价值。数据资源库共享着眼于数据资源的应用结果,是面向用户的用语,其价值与用户的目标相关。

20世纪80年代至90年代中期是商用数据库的鼎盛时期,大量商用数据库公司通过自己专用的远程计算机终端网络向用户提供数据库服务,美国的数据库公司最多,如路透、美联、彭博等。商用数据库公司主要提供金融行情、经济新闻、国际贸易、信用、法律、科技等信息,在信息资源稀缺的时代,商用数据库公司获得了很高的利润。商用数据库服务价格十分昂贵,信息服务对象主要是机构用户,特别是大型机构。20世纪90年代中期以后,廉价标准的互联网接入打破了数据库企业的专网壁垒,撤除了阻挡竞争者的高门槛,各种信息服务蜂拥而至,引发了过度竞争。

互联网技术的大普及带来了商用数据库市场的大萎缩,但是,机构内部使用的专用数据库反而借互联网技术的大普及之势获得了更大的发展,这是由于按经济学规律,供应增加会带来边际效益下降,过量的资料供应会带来信息资料的贬值,互联网时代的资料远不如数据库时代有价值。为了同时兼顾信息提取的成本和获取信息的效果,需要在一定范围内进行数据资源库的共享。例如,在基础数据资源库的前提条件下,建立人口、法人单位、自然资源和地理空间、宏观经济基础数据库四库合一型共享交换平台,在统一的标准规范下,建立包括数据抽取、反馈、纠错、复用等对各业务部门的业务具有良好支持的数据交换机制和信息资源目录与交换体系,实现对基础信息的统一管理、聚合和承载服务,满足全体公民对基础信息的共享交换的需求。支撑市民卡、数字医疗、综合治税、行政审批、数字城管、应急指挥、城市规划、辅助决策系统等典型数字城市应用。共享交换平台提供的服务包括海量栅矢数据的入库管理、发布与自动更新,多源异构政务信息资源的目录交换服务,后台强日志监控、强统计分析等;提供人口、法人等业务数据与地理空间信息的融合机制,具备强大的图层管理与共享交换的功能;适应“物理分散,逻辑集中”的建库原则以及多级部署、多种接入系统设计方式;提供标准服务接口,支撑搭建式、综合式、嵌入式、接口式、配置式等多种快速应用开发模式;提供各委办局业务数据的共享交换与管理机制,使其通过平台能快速搭建综合业务应用系统。

长期以来,数据资源库共享一直是一个难题。有些人会用保密问题为不愿共享辩解,其实保密不是障碍,因为数据资源库共享可以在同样的密级范围内推进,关键问题是各部门都不愿放弃部门数据的管理权,而多数数据资源库共享方案要求集中各部门数据,由某部门统一管理实现信息共享,在一些人的观念中数据不集中管理就无法实现共享。而谷歌、百度等企业认为数据资源的共享与数据的集中管理是两回事,搜索软件不需要数据集中管理也能实现资源共享。以高效搜索系统替代数据集中管理,创建共享数据资源库会更

经济有效。例如,很多地方政府建设企业信用系统,还是采取先集中各部门数据建立数据库再提供服务的老模式,其成功率很低,部门数据集中管理不仅行政阻力大,而且数据的更新和质量维护十分困难。但是,如果各部门按《中华人民共和国政府信息公开条例》的要求公开资料,由专用检索软件到各部门搜索一遍,将结果整合即可实现政府完整的企业信用信息服务,新模式不仅简单易行而且服务质量更好。

此外,共享数据资源库要求利用民间创意逐步开放数据和服务。到目前为止,我国政府的数据资源库并没有得到充分的利用,主要由于政府机制的制约和创意不足,只适合于完成规范化的任务。与政府相比,企业与社会有着近乎无限的创意,充分利用社会创意是政府信息服务的未来方向。目前在一些电子政务先进的国家和地区,政府已经开始将数据库资源更彻底地向公众与企业开放,充分利用民间丰富的创意资源,支持企业与非政府机构利用政府的数据库资源提供更有创意的增值服务,这也是共享数据资源库发展的方向。

3.1.5 特定数据资源库

特定数据资源库可被视为一种特殊的电子化文件柜——存储特定电子文件的处所,用户可以对文件中的数据进行新增、截取、更新、删除等操作。

在智慧城市建设过程中,特定的用户常常需要把某些特定的数据放进这样的“仓库”,并根据管理的需要进行相应的处理。例如,卫生服务、社会保障服务、国土资源监测、财务审计业务所涉及的特定数据资源库等。特定的数据资源库是为了满足特定使用者的目的而形成的,由于各个机构使用的系统不同,产生的文件格式也不同。因此,特定数据资源库具有以下特征:首先,它是非公开的;其次,数据库设立的目的是为了提供某种特定服务;第三,数据是经过分析、整合后录入的;第四,数据资源库的使用者是特定的、少数的。

特定数据资源库建设的起点是对部门的法定职责、用户的特定需求进行分析,确定信息化系统的发展目标、系统技术目标、信息化方式目标和信息化效能目标,构建顶层设计模型,在此基础上构建特定数据资源库架构。

典型的特定数据资源库如我国国家审计署建立的金审工程数据资源库,审计署围绕审计信息化近15年发展所达到的效能目标和业务发展情况,从“十五”时期关注差错纠错的信息效能目标,到“十一五”时期关注重要行业健康发展的信息效能目标,再到“十二五”时期主要关注国家经济安全和绩效审计目标的前期准备,展望了国家审计信息化今后一个时期的重点工作,即构建公共财政安全和绩效的审计数据资源库、构建公共财政安全和绩效的审计评价指标体系、开展公共财政安全和绩效的审计监督和评价。

国家审计履行法定职责的基本方式是,获取被审计单位的审计所需数据,转换成符合审计数据结构和格式要求的审计数据并建成审计数据库,开展数据分析,发现问题,提出审计意见和建议。在这一过程中,由于被审计单位的信息系统数据库表结构各异,而在审计机关之间需要信息共享,涉及采集被审计单位的数据接口标准、转换成审计所需的数据结

构和格式标准、审计数据库建设标准等三类数据标准^[10]。

为保障国家审计数据中心和省级地方分中心数据库的规范存储,审计署组织编制了《国家审计数据中心数据库建设规范》(计算机审计实务公告第32号)。规划的审计数据库体系包括元数据库、主数据库、审计信息资源目录库和各专业审计数据库。该数据库体系用于支撑国家审计数据中心的数据加载、数据分析、数据交换、模拟审计、电子沙盘和系统管理等应用。元数据是定义和描述其他数据的数据,规划的元数据分为技术元数据和业务元数据;主数据是指应用系统或者应用软件之间共同使用、具有稳定性的数据,规划的主数据包括审计机构与组织机构代码等;数据集市是用于分析相关专门业务问题或功能目标而做的专项数据集合。国家审计数据中心数据仓库按照各专业审计分类规划,包括预算执行审计、税收审计、海关审计、金融审计、企业审计、社保审计、固定资产投资审计、农业与资源环保审计、外资运用审计、经济责任审计、境外审计和其他专项审计数据集市。按照数据分析的业务需求,在专业审计数据集中或专业审计数据集市之间构建主题数据。《国家审计数据中心数据库建设规范》对审计信息资源目录库和专业审计数据库进行了规范,规划的审计信息资源目录库用于存储审计管理类和审计业务类信息资源目录的相关数据;规划的专业审计数据库包括物理存储的数据表和视图。其中,物理存储的数据表包含专业审计数据规划中的基础表和分析表,数据表及其字段使用专业设计数据规划确定的“标识符”,视图展现的数据库表及字段使用专业审计数据规划确定的“中文名称”。

又如上海万达信息国土支护监测平台数据资源库,该平台以万达信息基础平台为核心,结合上海等地实施经验,整合省、市、县三级国土资源行业信息(业务信息、基础信息),以数据仓库为载体,将RS、GIS与可视化和虚拟现实技术紧密地结合起来,侧重动态监测、实时监管并直观准确展示土地、矿产及地质环境管理的全生命周期数据信息,为领导决策和管理提供信息支持。其数据资源库的建设主要采用数据抽取整合的方式——将存在于各业务系统单线条的数据,抽取和整合到统一的数据仓库中,作为监测监管、展现分析、决策支持的基础。按照“纵向到底、横向到边”的要求汇集土地管理、地矿管理等多项业务板块和行政审批的全部信息,构筑国土数据仓库,为资源管理提供了“统一底板”。

值得一提的是,如果为了满足档案项目管理、数据库共享或公众查阅等的需要,特定数据资源库在具体应用中还需要根据以下6个方面进行修正性和适用性管理^[11]:

- (1) 提高公众使用在线系统访问数据库的能力,提高在线公布电子数据的灵活性,加强公众互动。
- (2) 促进数据提交过程的流程化,增强可扩展性、可靠性和使用的灵活性。
- (3) 提高工作人员检索和访问文件信息的基本能力。
- (4) 改善跨实例和跨系统捕获、存储和更新元数据的过程。
- (5) 改善信息系统架构以提高其功能的可扩展性和发展的可持续性,提高数据保存的成本效益,提高文件管理服务水平。
- (6) 根据智慧城市建设需求的特殊性,对数据资源库的体系和系统进行动态修正性维

护,提高数据资源库的开放程度。

3.1.6 公共服务数据资源库

1) 公共服务

所谓公共服务就是指使用公共权力和公共资源向公民(及其被监护的未成年子女等)所提供的各项服务。公共服务是有国家行为介入的一种服务活动。公共服务体现的是公民权利与国家责任之间的公共关系。公共服务可以使公民的某种直接需求得到满足,如教育和医疗保健。公共服务可以由公民根据个人需要进行一定程度的选择。公共服务涉及的人与人之间的关系是平等的。公立学校和公立医院等是专门的公共服务机构。例如,教育服务本身只是特定专业性服务,使用了公共权力或公共资源所提供的教育服务才是公共服务。而为了个人牟利使用私人资源所提供的教育服务或私立教育是营利性的私人服务。而非营利社会组织使用来自捐赠等渠道的社会资源所提供的教育服务或所办的公益性学校则是非营利性的社会公益性服务。

2) 公共服务的界定

在广义上,可以将公职人员使用公共权力与资源所从事的各项工作都看作是公共服务。当国家是建立在普遍的公民主权基础之上的时候,国家具有公共性质,国家存在的目的和职能就是为全体公民的利益和需求服务。在这个意义上,由国家的公共性质所决定,国家体系中的所有机构,如立法机构、行政机构和司法机构等都是提供公共服务的机构,在这些机构中任职的人们的工作都是在提供公共服务。在中国,这意味着在人大、法院、国务院以及各地方政府等国家机构中的工作人员都是在从事公共服务。但是,对公共服务的概念只有作狭义、具体和明确的界定才能在理论和实际操作上具有实质性意义,即提供公共服务是国家的主要职能之一,有其具体的内容和形式,并且可与政府或国家的其他职能相区分。也就是说,公共服务只是同经济调控、市场监管、社会管理相并列的国家的又一项职能。

狭义的公共服务不包括国家所从事的经济调节、市场监管、社会管理等一些职能活动,即凡属政府的行政管理行为,维护市场秩序和社会秩序的监管行为,以及影响宏观经济和社会整体的操作性行为,都不属于狭义公共服务,因为这些政府行为的共同点,是它们都不能使公民的某种具体的直接需求得到满足。公民作为人,有衣食住行、生存、生产、生活、发展和娱乐的需求。这些需求可以称作公民的直接需求。公共服务满足公民生活、生存与发展的某种直接需求,能使公民受益或享受。譬如,教育是公民及其被监护人,即他们的子女所需要的,他们可以从受教育中得到某种满足,并有助于他们的人生发展。如果教育过程中使用了公共权力或公共资源,那么就属于教育公共服务。

本章所指的公共服务的概念是狭义的,即能够满足公民直接需求的由国家介入的服务活动,如教育、医疗保健、社会保障以及生态环境保护等。

3) 公共服务的类别

公共服务可以根据其内容和形式分为基础公共服务、经济公共服务、社会公共服务和公共安全服务。基础公共服务是指通过国家权力介入或公共资源投入,为公民及其组织提供从事生产、生活、发展和娱乐等活动都需要的基础性服务,如提供水、电、气,交通与通信基础设施,邮电与气象服务等。经济公共服务是指通过国家权力介入或公共资源投入为公民及其组织即企业从事经济发展活动所提供的各种服务,如科技推广、咨询服务以及政策性信贷等。社会公共服务则是指通过国家权力介入或公共资源投入为满足公民的社会发展活动的直接需要所提供的服务。社会发展领域包括教育、科学普及、医疗卫生、社会保障以及环境保护等领域。社会公共服务是为满足公民的生存、生活、发展等社会性直接需求,如公办教育、公办医疗、公办社会福利等。公共安全服务是指通过国家权力介入或公共资源投入为公民提供的安全服务,如军队、警察和消防等方面的服务。

4) 公共服务数据资源库

公共服务数据资源库包括四大类:基础公共服务数据库、经济公共服务数据库、社会公共服务数据库和公共安全服务数据库。

(1) 基础公共服务数据库包括:水、电、气,交通与通信基础设施,邮电与气象服务等数据库。

水文数据库:水是生命之源、生产之要、生态之基,水文数据是经济社会发展不可替代的基础支撑和保障,水文工作具有极强的公益性、基础性、战略性。水文数据库建设不仅事关水文、水利事业发展,而且事关经济社会发展全局;不仅关系到防洪安全、供水安全、粮食安全,而且关系到经济安全、生态安全、国家安全。

国家水文数据库是1986年纳入国家计委、国家经济信息中心制定的《国家级数据库群建设方案》中第一批优先开发的72个数据库之一。福建省国家水文数据库系统作为国家水文数据库的福建节点,承担着福建省全部水文站点自有刊印年鉴以来的所有水文年鉴数据的计算机管理工作。在建设过程中严格执行由原水利部水文司、水利信息中心制定的各项技术规范,并以满足水利部提出的国家水文数据库基本建成的技术标准作为工作目标,保证了福建省国家水文数据库的标准化和规范化。以福建省水文数据库建设为例,福建省水文数据库建设工作,大体可分为3个阶段:准备及初试阶段(1988~1994年);录入阶段(1995~1999年);建库阶段(2000~2001年),2001年完成验收。数据库资料截止到2011年,共有水位6664站年、流量3992站年、沙1834站年、潮水696站年、雨量39024站年、蒸发2967站年。数据库3.0版数据文件868MB,4.0版共5120MB。

水利部高度重视水文基础工作,部署要求加快推进国家水文数据库系统建设。水利部水文局于2009年组织完成了《国家水文数据建设规划》并通过专家审查,在此基础上将主要内容汇入了《全国水文基础设施建设“十二五”规划》,力争在“十二五”期间,建成基于网络、集中与分布相结合的国家水文数据库系统,完成全国范围核心水文数据的实体装载,省级以上节点具备基本再发展能力和服务能力。

城市用水数据库包括:降水量管理、地表水管理、地下水管理、水源管理、水环境管理、

需水与用水管理、供水管理、水质管理、取水许可证管理、水费征收管理数据库。城市用水数据库又分为空间数据和属性数据,空间数据是用来反映研究区空间特征的要素,如地形、位置、形状、面积、水系、土地使用、土壤、地表特征等以及它们之间的拓扑关系;属性数据信息用来反映与地理实体相联系的地理变量或地理意义,如需水量、供水量、水资源监测数据、地区的经济与人口状况、行政分区、河流和排污口的基本情况、废水的排放量及污染物的种类等与水质有关的基本数据等。

国家水文数据库和城市用水数据库,可为城市用水管理、水资源公报查询、水利工程信息查询、水资源宏观决策等提供强大的水资源数据管理、评价计算、统计查询等功能,大大提高城市水资源管理的能力和科学化、现代化水平,增强水资源主管部门的协调能力和决策能力,提高工作效率,为城市的水资源规划、管理和决策提供及时、准确的信息服务和先进的技术支持。

电力数据库包括:空间电力数据库、城市配电网海量设备信息库、城乡居民生活用电数据库等。城乡居民生活用电数据库包括乡村和城市用电数据库。全行业用电数据库包括第一产业、第二产业和第三产业用电数据库。分行业用电数据库包括农、林、牧、渔、水利业,排灌,工业(轻工业、重工业、制造业),建筑业,交通运输、仓储、邮政业,信息传输、计算机服务和软件业,商业、住宿和餐饮业,金融、房地产、商务及居民服务业,公共事业及管理组织用电数据库。

城市智能燃气网的数据库包括:输配管网地理信息数据库、数字管道数据库、风险预评价数据库、用户需求计量数据库、管网运营智能化的监控数据库等。利用这些数据库,采用云计算技术,为我国城市输配管网数据采集与监控系统(SCADA系统)、地理信息系统(GIS系统)、企业资产管理系统(EAM系统)、管网运行系统、模拟仿真系统等诸多与燃气管网建设运营相关的子系统提供数据支持。提高系统的分析处理能力,为城市智能燃气网的建设发展提供信息保障。

智慧交通数据库包括:交通实况数据库、视频监控数据库、交通诱导数据库、电子警察数据库、智能卡口数据库、道路交通管理数据库(车辆管理、驾驶员管理、违章管理、事故处理)、车辆信息数据库、停车场信息数据库等。利用这些数据库信息,使各种动、静态信息浑然一体、相互补充,便于指挥人员迅速决策、快速反应与处警,使广大交通出行者全面掌握监控区域的交通状况,及时修正交通计划,保证交通的安全与畅通。

(2) 经济公共服务数据库:从事经济发展活动而提供的各种服务所需要的数据库。如,国家和城市宏观经济数据库、城市各种生产生活品价格数据库和城市各种生产生活品市场需求数据库、科技创新与知识产权数据库、政策性信贷数据库等。

(3) 社会公共服务数据库:公共教育数据库、公共科普数据库、公共医疗卫生数据库、公共社会保障数据库和城市环境数据库等。

(4) 公共安全服务数据库:国防安全数据库、事故灾害数据库、地质灾害数据库、地震防控数据库、突发事件及预案数据库、食品药品安全数据库等。

3.1.7 管理数据资源库

这里所指的管理为公共管理,公共管理是以政府为核心的公共部门整合社会的各种力量,广泛运用政治的、经济的、管理的、法律的方法,强化政府的治理能力,提升政府绩效和公共服务品质,从而实现公共的福利与公共利益。公共管理的主体包括社会公共组织和社会其他组织两大类。公共管理的客体是社会公共事务,社会公共事务具体分为公共资源和公共项目、社会问题等。公共管理的目的是推进社会整体协调发展和增进社会公共利益实现,其职能是调节和控制。

管理数据资源库包括:城市基础地理数据库、公共设施数据库、公共产品数据库、公共信息资源数据库、人力资源数据库、自然资源数据库、公共项目数据库、社会问题数据库。其中,公共信息资源数据库包括:公共事业数据库、公共政策数据库、文化产业数据库、体育产业数据库、科技成果数据库、经济信息数据库。人力资源数据库包括:城市人口数据库、劳动关系数据库、劳动与社会保障数据库、教育数据库、市民健康数据库等。自然资源数据库包括:土地资源、矿产资源、林业资源数据库,渔业资源数据库和航运资源数据库等。社会问题数据库包括:公共安全管理数据库、应急管理数据库、国防教育与管理数据库、食品经济管理数据库、海关管理数据库等。

3.1.8 政府决策数据资源库

政府决策主要分为宏观规划决策和专项政策决策两大类。

1) 宏观规划决策及数据资源库

(1) 宏观规划决策。宏观规划必须体现中央、国家的经济政策,以市场为依据,优化资源配置。规划目标必须合理,措施必须可行,预测力求科学并接近实际,各产品、业务之间必须协调发展。工信部是我国信息产业宏观规划的负责部门,在制定规划的具体过程中,要和国务院及其下属的国家发展计划委员会和商务部密切配合。宏观规划制定的具体程序如下:

- ① 接受任务,根据中央国务院要求开始规划工作,一般情况是按照国民经济发展五年计划,每五年编制一次。
- ② 确定目标和总体框架,决定编制方式(元决策)。
- ③ 分解目标,各司局根据职能分工编制本专业的规划。
- ④ 将分专业规划汇总、平衡,协调产业产品间的关系,地区平衡。
- ⑤ 拟订规划预案。
- ⑥ 征求意见。
- ⑦ 根据征求意见情况修改后报分管副部长审阅。
- ⑧ 部长办公会审定批准,由国家发展计划委员会汇总综合制定国家总体规划。

⑨ 国务院批准后,由工信部对外发布实施。

决策过程中的信息传递主要是上下层面之间的直线传递。不同决策层面对信息的需求差异很大,从国家宏观经济政策,直至某一类具体产品的市场需求。

各决策层面对总体目标的认识层次是不同的,规划的着眼点也是不同的。总目标力求总体的发展和行业、地区之间协调平衡,而专业规划的目标希望本专业得到更多的优惠,都会强调各自的重要。不过这两方面比较容易协调,一是因为两者都站在国家的高度,没有利益上的冲突。二是同处在一个政府部门,信息容易沟通取得一致,而地方规划和企业规划的出发点是完全不一样的。地方政府首先考虑的是本地经济发展问题,各地均从自己的利益出发进行规划,企业也是如此。这无可非议,是其职责所在,是市场经济中最合理的行为。但这引申出一个问题,如何客观公平、公正地平衡地方与企业的规划?而从实际中看这是决策过程中的一个薄弱环节,政府决策数据资源库能针对这一问题提供极大的帮助。

大量的决策是在司处一级完成的。实际上,最终的决策人只是对一系列决策的结果构成的框架和数据及数据间的关系进行认可,他不会也不可能去研究报上来的整个规划。政府决策数据资源库能对这一状况提供极大的帮助。

预测在规划中的作用十分重要,许多决策都是以预测的结果为依据的,依据政府决策数据资源库使预测更准确、更科学,避免将假设当成预测。

专家和专家论证会是规划决策的基本支持模式,多数政府机关将其作为科学决策的主要手段,专家的依据就是政府决策数据资源库。分析发现,在综合汇总过程中,比较判断的工作量很大,平衡的方式、制定的标准对决策结果影响很大。以政府决策数据资源库为依据,将使决策结果更加客观。

(2) 宏观规划决策数据资源库。政府决策数据资源库是政府宏观规划决策的必要支撑。宏观规划决策数据资源库包括:国家宏观经济数据库、国家宏观经济政策数据库、宏观市场需求数据库、国家自然资源数据库、行业规划数据库、行业预测数据库、产业规划数据库、产业预测数据库、产品与产业关系数据库、专业规划数据库、地区经济数据库、地区资源数据库、地区规划数据库、地区平衡数据库、规划预案数据库、企业规划数据库等。

2) 专项政策决策及数据资源库

(1) 专项政策决策。每个产业的专项产业政策是针对每个产业的发展情况,调节产业内部的结构比例、区域分布,促进关键产业、产品的发展。根据产业政策调节的对象不同,由不同的司局负责制定,产业政策制定程序如下:确定需要进行特殊调控的对象,主要依据是市场反映出的情况和中央、国务院的有关政策指示;决定承办单位,明确工作内容;组织专项调研和专家讨论;确定发展目标;拟订政策措施;部长办公会议审议;向政策内容所涉及的相关部委正式征求意见;上报国务院审批后颁布实施。

目前某个产业所涉及的产业政策,均是着眼于扶植关键产业的发展,这是由行业的特点所决定的。对于通信行业,属政府管制经济的范围,即使是放开经营的业务,也需向政府申请许可,政府可采用更加直接的手段进行调控。对于电子信息产品制造业,由于是放开

经营的市场竞争体制,实行的是宏观规划指导下的市场竞争,其结构和组织区域布局均主要靠市场调整。事实上,政府曾试图利用规划布点等手段对某些产品进行调整,最突出的例子是彩电。因而目前重点是加强对瓶颈产业的支持力度,扶持新的经济增长点。需扶持的产业一般属基础范畴,投入高,技术难度大,资本流动差,投资风险大,如集成电路、新型显示器件等。另外软件由于其特殊性也被作为基础产业。

鉴于目前国家的管理体制,政策涉及的金融、财税、工商及进出口等政策都由相应的部门负责。商务部、工信部作为负责制定信息产业有关产业政策的部门,先提出建议,征得相应的部门同意才可实施,否则要么取消要么交国务院裁决,因而一个产业政策要经许多部门,甚费时间,有可能最后胎死腹中。

产业政策制定的基本方法,一般由政府部门与行业支持咨询机构及专家共同组成小组。通过调研访问,摸清有关产业的发展状况、存在的问题,剖析问题的成因,针对问题提出解决方案,综合各种方案形成政策草案,征求意见后上报审批。研究过程和各阶段的决策支持,需要专项政策决策数据资源库。

产业政策措施一般是宏观微观并用,但对政策手段的调控力度缺乏深入研究,因而在实施过程中,出现措施无效的情况,要加强措施针对性。

产业政策研究中多依赖专家的意见,专家多是来自政策要调控的行业,需保证专家意见的科学性、全面性,专项政策决策数据资源库可为其提供科学支撑。

征求其他部门意见时经常会通不过。一个原因是对措施的效果用定性的描述多,深入的分析少,对政策的负面效应分析少,没有政策成本的概念,不能说服人;另一个问题是政策措施涉及其他复杂的问题,可操作性差,因此也需要专项政策决策数据资源库提供支撑。

(2) 专项政策决策数据资源库。包括:国家宏观经济数据库、国家微观经济数据库、产业政策数据库、产业结构数据库、产业比例数据库、产业区域分布数据库、关键产业数据库、产品发展数据库、微观市场数据库、行业数据库、金融数据库、财税数据库、工商数据库、进出口数据库、产业发展状况数据库、产业问题数据库、产业问题成因数据库、产业解决方案数据库等。

3.2 智慧城市数据资源交换体系和目录体系

智慧城市数据资源交换和目录体系是建立信息资源物理分散、逻辑集中的信息共享模式,提供一定范围内跨部门和组织、跨地区的普遍信息共享,方便用户发现、定位和共享多种形态的信息资源,支持政府的经济调节、市场监管、社会管理和公共服务。支持政府、医院、学校、企业、市民之间横向按需信息交换与共享,支持不同异构应用系统间信息交换与共享,实现信息资源的高可靠性、可用性。

基于目录体系的智慧城市数据资源整合具有业务驱动、按需整合、满足信息整合整体

性原则及标准化的优势,可以支持智慧城市数据资源整合和共享需求,同时可以适应数据资源随业务动态变化的特点,保证整合的数据资源及时准确,是实现智慧城市数据资源整合和共享的先进手段,是智慧城市建设的基础设施。

3.2.1 数据资源交换体系

1) 数据资源交换体系建设需求

数据资源交换体系建设是推动智慧城市管理信息资源共享服务一体化建设的重要工程技术手段,将有效支持智慧城市推广和建设,基于项目建设以及实际业务运行环境、流程沉淀的一系列技术标准规范,形成智慧城市数据资源共享长效机制等目标。

完成基于智慧城市数据交换体系的平台建设,将提供数据共享交换服务、交换服务组件配置管理、平台运行管理、日志统计分析等服务,并基于强大的可扩展性、自适应性、对异构系统的兼容性,支撑智慧城市各类信息资源的共享交换和各应用系统之间的互联互通,促进智慧城市中不同组织不同业务协同协作。

智慧城市数据资源交换体系是依托智慧城市信息安全基础设施,为跨部门和组织、跨地域数据资源交换与共享提供的数据服务基础设施。交换体系总体结构由服务模式、交换平台、信息资源、技术标准和管理机制组成,如图 3-4 所示。不同服务模式的业务应用通过

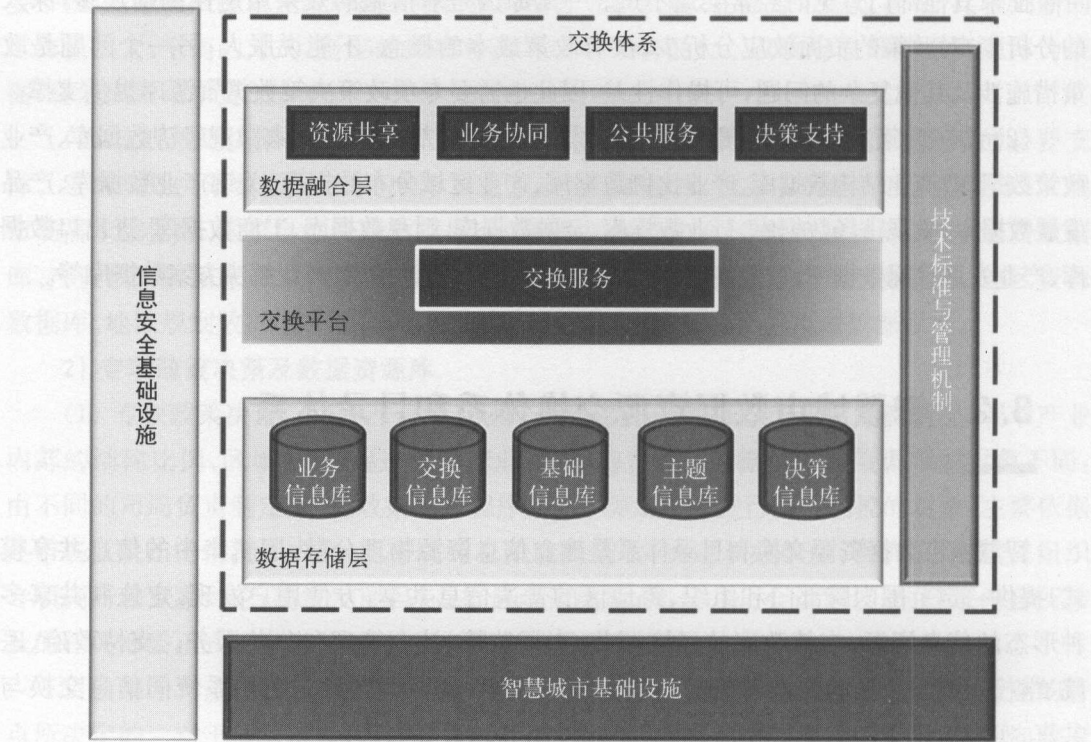


图 3-4 智慧城市数据资源交换体系总体结构

调用交换平台提供的交换服务,实现对信息资源的访问和操作,技术标准和管理机制为数据资源的交换和共享提供技术和管理的保障^[12]。

2) 数据资源交换概念模型

交换概念模型如图 3-5 所示,由中心交换节点和端交换节点组成。端交换节点接收和发送各组织的交换信息。中心交换节点管理交换域内端交换节点的数据交换服务,并根据需求形成共享信息库。

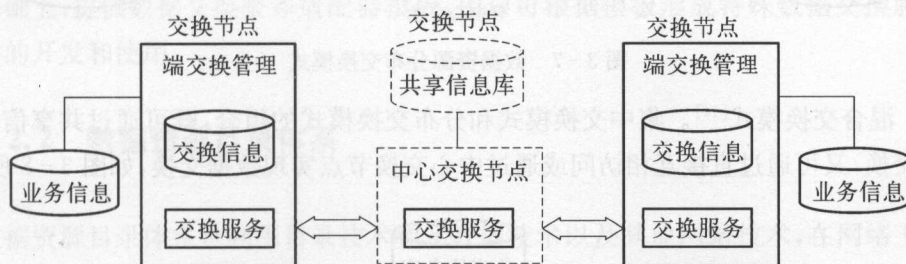


图 3-5 数据资源交换概念模型

- (1) 业务信息是由各组织产生和管理的智慧城市数据资源。
- (2) 交换信息是端交换节点用于存储参与交换的智慧城市数据资源。
- (3) 共享信息库是可以为多个端交换节点提供一致的智慧城市数据资源的信息集中存储区。任意一个端交换节点可以按照一定的规则访问共享信息库。在交换体系中,共享信息库是可选的。
- (4) 端交换节点是智慧城市数据资源交换的起点或终点,完成业务信息与交换信息之间的转换操作,并通过交换服务实现智慧城市数据资源的传送和处理^[13]。
- (5) 中心交换节点主要为交换信息提供点到点、点到多点的信息路由、信息可靠传送等功能。在两个端交换节点之间可以有若干个中心交换节点。
- (6) 交换服务是交换节点传送和处理智慧城市数据资源的操作集合,通过不同交换服务的组合支持不同的服务模式。

3) 数据资源的交换模式^[13]

(1) 集中交换模式。数据资源集中存储于共享信息库中,信息资源提供者或使用者通过访问共享信息库实现信息资源交换,如图 3-6 所示。

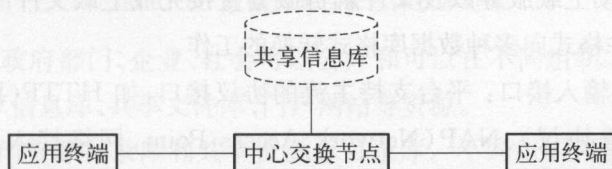


图 3-6 数据资源集中交换模式

(2) 分布交换模式。数据资源分布存储于各业务信息库中,数据资源提供者 and 使用者

通过交换节点提供的交换服务实现两者之间信息资源定向传送的交换模式。分布式交换模式可划分为有中心交换和无中心交换,如图 3-7 所示。

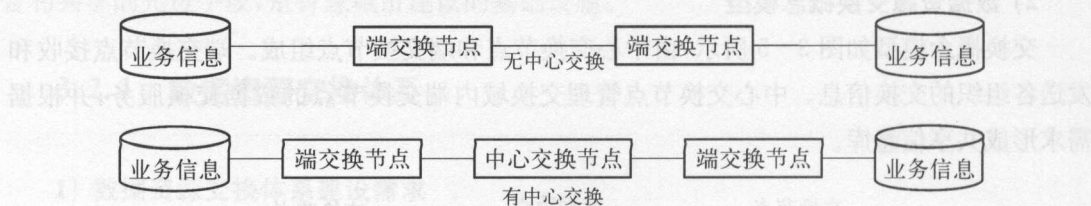


图 3-7 数据资源分布交换模式

(3) 混合交换模式^[14]。集中交换模式和分布交换模式的组合,既可通过共享信息库实现信息交换,又可通过直接互相访问或通过中心交换节点实现数据交换,如图 3-8 所示。

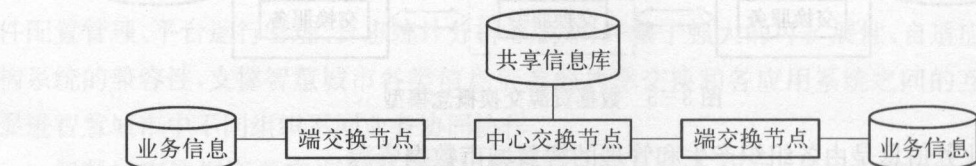


图 3-8 数据资源混合交换模式

4) 四类节点的接入方式

根据智慧城市建设要求,数据交换平台将根据实际业务环境,提供前置交换节点、目录管理节点、子平台节点和应用系统节点四种接入方式,便于对各类共享服务进行有效推广和应用。

5) 数据交换平台接口要求

数据交换平台应能够提供数据库、文件系统等多种接口方式,提供如下输入、输出接口:

(1) 数据库级输入接口:平台支持各类主流关系型数据库的接口,包括 Oracle、SQLServer、Sybase、DB2 等,可以通过数据库适配器直接访问各个业务系统的数据库,从而实现数据同步和数据通信的目的。

(2) 文件级输入接口:平台支撑各类主流文件类型接口,包括 Excel、TXT、DOC、PDF 等,可以通过文件自动上载服务以及文件解析服务直接完成上载文件的交换,同时根据目标节点需求完成文件格式向多种数据库格式转换的工作。

(3) 协议级交换输入接口:平台支持主流的协议接口,如 HTTP(Hypertext Transfer Protocol,超文本传输协议)、NAP(Network Access Point,网络接入点)、Socket、SOAP(Simple Object Access Protocol,简单对象访问协议)等协议接口,向业务系统发送请求或接收响应。

(4) Web 服务输出接口:对于平台本身提供的 Web 服务,或其他各单位提供的 Web

服务,应用方可直接调用其接口。

(5) 非 Web 服务输出接口: 对于非 Web 服务的行业应用系统接口,共享交换平台应提供统一、方便的接口技术,实现应用系统与共享交换平台的结合。

(6) GIS(Geographic Information System,地理信息系统)服务接口: 提供城市 GIS 数据的存储区域,并完成相关应用系统调用 GIS 数据的接口。

(7) 交换服务适配器模板服务接口: 数据交换平台建设完成后,将在以上各类通用接口的基础上,提供数据交换服务适配器模板,用户可根据模板形成特殊数据交换服务适配器组件的开发和使用。

3.2.2 数据资源目录体系

数据资源目录体系是利用目录技术和元数据技术以及其他网络技术,在网络上构造一个统一的智慧城市级别数据资源目录管理系统,使资源的利用者能够在任何时间、任何地点,通过特定的服务接口查询资源目录,使其能够快速发现、定位和获取所需信息。

1) 目录体系总体技术框架

数据资源目录体系是向信息使用者提供政务信息查询、检索和定位的服务平台,并在规定的安全机制下,通过交换体系获得信息资源,向信息使用者提供信息访问服务,如图 3-9 所示。

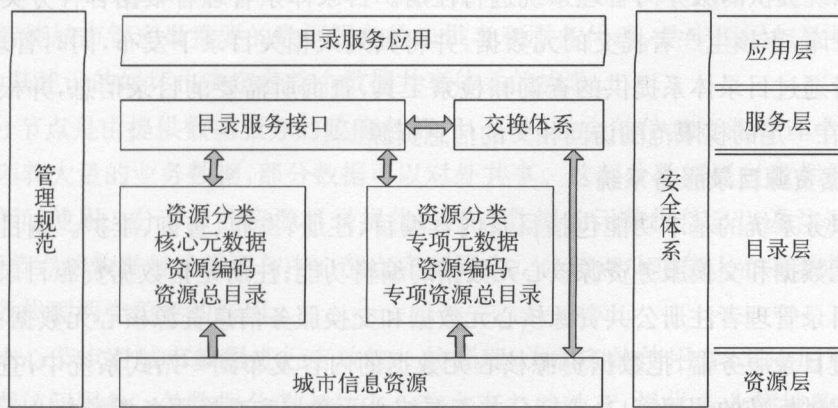


图 3-9 数据资源目录总体技术框架

资源层是各级政府部门、企业、社会可以公开和可以在不同组织之间共享的政务信息资源,包括各类共享信息库、共享文件库、门户网站等资源。

目录层包括专门资源目录库和共享资源总目录库。不同组织可以根据协同应用的需要,建立组织之间共享指标项目目录库^[15];根据对公共服务应用的需要,建立门户网站服务目录库;根据本领域应用的特点,建立相应的专项资源目录库。随着专项资源目录库建设的不断成熟,共享资源总目录库也逐渐形成。

服务层主要包括目录体系向应用层或其他系统提供各类应用服务的接口,以方便用户的调用、目录体系与交换体系互通、目录体系之间的信息交换和访问。

应用层是目录服务向用户的展示层。用户使用应用层提供的各类工具进行高信息资源的检索、查询、访问,也可以进行信息资源的著录和注册,以及对目录库进行管理^[16]。

2) 目录体系服务模型^[15]

目录体系主要包括三个部分,即目录体系生产系统、目录体系管理系统和目录体系查询系统,如图3-10所示。其中使用者包括三类用户:元数据生产者、目录体系管理者和信息查询者。

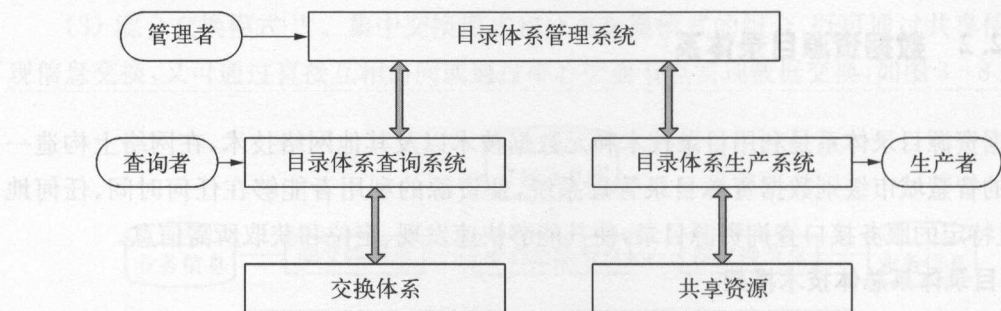


图 3-10 目录体系服务模型

生产者使用生产系统对本单位需要共享的信息资源按照相关的元数据标进行著录,并通过管理系统提供的服务向管理系统进行注册。目录体系管理者根据各种分类体系构建相关的目录库,审核生产者提交的元数据,并将其列入相关目录下发布,同时管理目录库。信息查询者通过目录体系提供的查询和检索工具,查询所需要的目录信息,并根据目录信息的指引,在一定的权限范围访问相关的信息资源^[17]。

3) 数据资源目录服务系统

目录服务系统的基本功能包括目录内容编目、注册、发布、查询、维护。编目提供公共资源核心元数据和交换服务资源核心元数据的编辑功能;注册是指数据资源目录提供者向数据资源目录管理者注册公共资源核心元数据和交换服务信息资源核心元数据;发布是指管理者通过目录服务器,把数据资源核心元数据的内容发布到一站式系统中;查询是指为应用系统提供标准的调用接口,支持公共资源核心元数据和交换服务资源核心元数据的查询。数据资源目录服务系统工作流程如图3-11所示^[12]。

智慧城市管理中的数据具有天然的分布性,各数据资源节点是分布的,应用也是分布的。共享服务平台提供的是一个逻辑统一的视图,其支撑是分级、多节点的数据与目录服务体系。具体来说,在一个城市内,可以建立城市数据分级、多节点共享体系,形成分布式数据管理、集中式服务发布,最终形成一个跨部门、分布式、可运行的数据服务与应用系统。提供数据与目录服务的是数据中心节点和分节点^[18]。

(1) 中心节点是智能城市管理跨部门的数据交换、共享的核心,通常其职能由城市的信

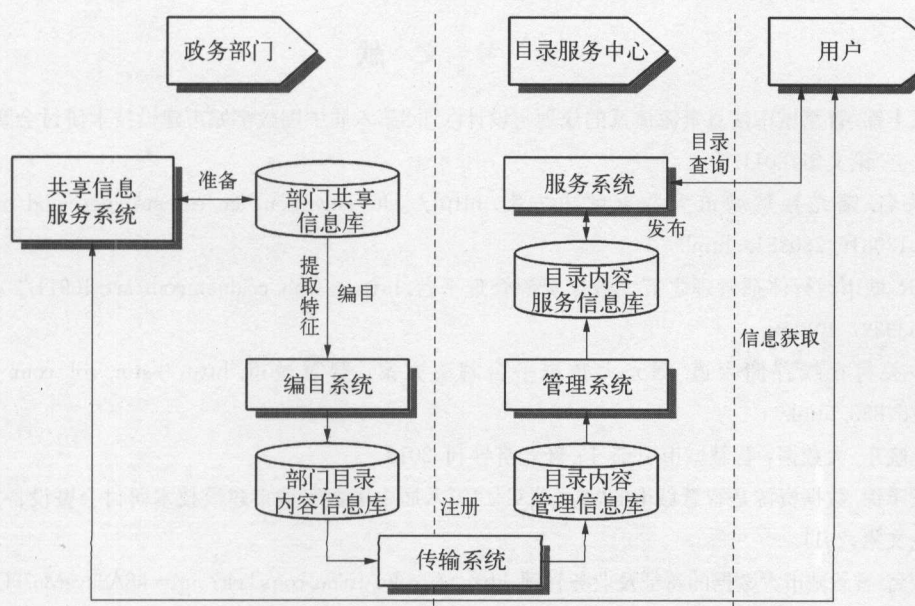


图 3-11 数据资源目录服务系统工作流程

息(资源)管理中心完成。一方面中心节点管理着城市中绝大部分数据,可以在城市层面进行统一规划,开发数据服务器,提供支撑不同委办局应用系统的服务接口,实现服务共享和互操作。另一方面在中心节点建立城市数据资源的总目录,对中心节点与分节点的数据资源进行统一的导航、检索和定位,这样必须确保整个中心节点的目录是不断更新的。中心节点作为智能城市管理最重要的数据服务的提供者和服务代理,是应用层与数据层中最重要的纽带,其建设的好坏也决定着整个数据共享体系的成败。

(2) 分节点是由提供数据服务的政府各部门、企业、社会的信息资源管理中心组成,各分节点管理着大量的业务数据,部分数据可以对外共享。这部分数据通过数据服务接口方式供其他部门使用。分节点的资源目录提供各节点数据的元数据信息以及注册的数据服务信息,分节点的数据目录需要与中心节点实现同步。分节点从功能上亦有数据服务的提供者和服务代理两类角色的作用。

数据中心节点对城市数据进行有效统一的管理,为服务的使用者,即应用系统提供绝大部分数据访问服务。各数据分节点是中心节点有益的补充,提供专业领域数据的服务。用户访问的入口可以是中心节点或分节点的数据资源目录,但对用户提供一个统一访问的一站式数据资源总目录更为理想,通过中心节点和分节点资源目录注册的同步实现。这样,应用系统只需通过对中心节点资源总目录的查询,就可以对中心节点和分节点中数据服务定位。应用系统所需的服务可能部署在中心节点,也可能在分节点,应用系统根据其服务质量进行选择^[18]。

◇参 考 文 献◇

- [1] 王丰锦. 智慧城市信息资源体系的规划与设计[C]. “第六届中国数字城市建设技术研讨会暨设备博览会”论文集, 2011.
- [2] 佚名. 曙光智慧城市大数据解决方案. http://edu.cnw.com.cn/edu-storage/cloud/htm2013/20130816_280381.shtml.
- [3] XR. 城市运行体征管理之五大数据资源管理平台. http://tech.ccidnet.com/art/40911/20130626/5032227_1.html.
- [4] 中关村在线存储频道. 曙光大数据平台解决方案: 智慧城市. <http://stor.zol.com.cn/375/3757833.html>.
- [5] 阮晓东. 大数据: 智慧城市引擎[J]. 新经济导刊, 2013.
- [6] 王丰锦. 数据资源是智慧城市的“基石”[C]. “第六届中国数字城市建设技术研讨会暨设备博览会”论文集, 2011.
- [7] 佚名. 智慧城市大数据的特征及业务管理. http://wenku.baidu.com/link?url=k8AdkygMp7FOKNQZe-KYMOk5GUzbt6l6uky_Mmgte6vRQRUzo2GrktEa9YSd9qN5Aq1dlsGoPyCzCOvdCmfoSzOGAIJ_rtjuTdDhxvgBVXq.
- [8] IBM. 云存储基础架构剖析. <http://www.uml.org.cn/yunjisuan/201106034.asp>.
- [9] 胡小明. 信息资源观念的变迁[J]. 电子政务, 2012(1).
- [10] 周德铭. 国家审计信息化的信息共享与业务协同[J]. 电子政务, 2013(7).
- [11] 张朋柱, 刁石京. 我国政府宏观决策任务的分类研究[J]. 系统工程学报, 2001(5): 354 - 459.
- [12] 韩霖. 政务信息资源共享交换平台设计与实现[D]. 北京: 北京邮电大学, 2010.
- [13] 宋晓虹. 民政低保信息系统总体框架研究及数据交换技术实现[D]. 长沙: 湖南大学, 2009.
- [14] 杨同建. 农业信息化服务平台的研究与开发[D]. 济南: 山东大学, 2012.
- [15] 牛燕. 一种电子政务目录体系中服务发现机理的研究与设计[D]. 长沙: 湖南大学, 2009.
- [16] 郭家义. 政务信息资源目录体系的相关问题初探[J]. 电子政务, 2005.
- [17] 佚名. 政务信息资源目录体系的总体架构. <http://www.e-gov.org.cn/news/news007/2008-09-17/94540.html>.
- [18] 林俞先. 数字城市中遥感数据共享服务的研究[J]. 计算机科学, 2007.

第4章

大数据在智慧城市应用中的关键技术

本章主要内容包括：数据生命周期管理、大数据采集、大数据存储、大数据处理、智慧城市大数据决策分析等。

其中，数据生命周期管理主要针对大数据场景下数据的共享性、完整性、可靠性和有效性的各个方面进行描述。其中引入了主数据管理、元数据管理、数据清洗及数据存储策略等各个方面。

大数据采集是针对如何在智慧城市场景中获取大数据的方法，其中主要介绍网络数据获取、媒体流获取、日志信息获取和传统应用数据获取。

大数据存储则主要介绍各种大数据的存储技术及硬件架构的选择，特别提出了基于CAP原则的各种存储技术选择，并以拓展的视野介绍了大数据存储的未来。

大数据处理主要介绍的是基于大数据方案的计算技术、数据挖掘技术和大数据可视化技术。

智慧城市大数据决策分析主要介绍融合多领域信息的数据知识模型，面向智慧城市决策支持的分布式海量数据的管理技术，辅助决策模型的智能组合技术和高性能多用户协同的辅助决策技术。

4.1 数据生命周期管理

数据生命周期管理主要是对智慧城市业务数据的产生、规范、使用和存档的各个阶段进行有效的监控和管理，主要体现在保证数据的共享性、数据的完整性、数据的可靠性和数据的有效性等各个方面，并以此为基础提供给上层决策分析系统。

(1) 数据共享性(主数据管理)：数据在其职能范围和组织范围内应实现共享，智慧城市应用网络内的用户可以无障碍地获取所需的数据来完成他们的工作，从而达到消除信息孤岛的目标。数据可以及时更新，保证各业务应用获取的数据是及时准确的，从而确保决策系统的有效性。

(2) 数据完整性(元数据管理)：各个智慧城市的业务单元信息资源将统一整合，消除数据的不一致，去除信息冗余，从而保证数据的完整性和唯一性。

(3) 数据可靠性(数据清洗)：智慧城市如此繁杂的现有系统中的数据存在很多问题，极易容易造成脏数据，可能的原因如下：滥用缩写词、惯用语不规范、数据输入错误、重复记录、丢失值、不同的计量单位和过时的编码等。数据必须进行有效清洗，确保数据来源以及传输的可靠性和可追踪性。

(4) 数据的有效性(数据存储策略): 智慧城市拥有海量的数据,而每次需要分析的数据可能只是其中的一部分,要做到有效地利用数据,需要让使用频率高的数据常驻内存,将一些使用频率低的数据做存档备份。

4.1.1 数据共享性: 主数据管理

智慧城市各个业务单元之间的信息系统呈紧耦合状态,各个信息系统之间存在资源冗余、重复建设、数据缺少整体规范、数据分散在各自应用系统中、业务中部分历史数据残缺不全等问题。因此,迫切地需要建立集中的主数据存储,以达到以下目的:

(1) 智慧城市的管理需要整体效应,加强各个业务单元的管理信息系统和应用系统之间的连接。

(2) 避免数据来源不一、数据统计和口径不统一,消除信息孤岛,为智慧城市利用各种数据进行深层分析和辅助决策提供数据基础。

(3) 降低软硬件投资和系统维护的总体成本,大幅提高基础设施的综合利用率,简化管理维护工作(存储备份、容灾切换等),提高系统和数据的安全性。

4.1.2 数据完整性: 元数据管理

信息资源是指在智慧城市运行、生产及管理过程中所涉及的一切文件、资料、图表和数据等信息的总称。它涉及智慧城市运行、生产和管理活动过程中所产生、获取、处理、存储、传输和使用的一切信息资源。这需要信息系统在智慧城市各个业务单元内提供数据采集、数据服务,规范数据流程管理、整合数据源头、统一数据产生源、统一数据格式。因而信息资源规划将定义信息资源的主要类别、信息资源规范、信息资源目录、信息资源分布存储和信息资源安全,以支撑整个智慧城市业务系统的正常运转,它不关心逻辑或物理的存储系统的设计。

信息资源规范是按照数据元标准化的相关原则和方法,对智慧城市的业务信息进行科学分类与编码来形成标准化文件。它有助于规范业务信息管理并指导信息化建设,是构建统一信息系统、实现信息资源管理一体化的重要前提。统一的数据资源是可扩展的信息资源(数据)的描述模型;提出信息资源规范编制的原则、内容和要求;定义智慧城市主要业务单元数据的命名、语法、语义和使用方法,从而实现整个智慧城市业务数据共享的目的。

信息资源规范主要内容包括数据元目录、代码集、维护与管理三部分。数据元目录是以目录形式排列的业务数据元的集合,反映业务管理中的基本信息对象的特征、行为、状态等内容。代码集的主要内容是代码表,代码表是业务数据元的取值范围,一般与相应的数据元配合使用。维护与管理是根据业务管理实际需要,按照标准化的要求,为数据标准的注册、更新与管理建立规则、方法与机制。

4.1.3 数据可靠性：数据清洗

脏数据可能带来如操作费用昂贵、决策制定失败甚至于错误、组织和下属的不信任、分散管理的注意力等后果,因此针对脏数据的有效处理是进行辅助决策的必要步骤。而数据清洗处理可以有效地清除脏数据、保证数据的质量。

数据清洗处理是构建数据仓库的第一步,鉴于数据的海量特性,不可能进行人工处理,因此自动化数据清洗受到工业界的广泛关注。提出了自动数据清洗的概念框架:定义和判断错误类型;查找并标示错误实例;修改没有发现的错误。

4.1.4 数据有效性：数据存储策略

智慧城市业务单元每天会通过物联系统、业务系统、移动应用产生极大量的数据。面对来源广泛、数量巨大、形式多样、内容丰富的大数据。随着智慧城市业务发展以及各种平台、系统的建立,也意味着需要管理的数据量越来越庞大,资源维护成本越来越高,效率越来越低。

面对智慧城市管理分析业务对数据的不同时效、不同访问频率、不同重要性等要求,如何判定数据是否有维护的必要性,如何降低数据维护成本、提高数据服务水平是进行数据生命周期管理需要解决的问题。

数据生命周期管理框架由数据归类、数据特性分析与数据存储策略三部分组成。数据存储框架首先对数据进行归类,在数据归类的基础上结合业务与系统实际情况,分析数据特性,最后根据现状调研、数据归类与数据特性制定数据生命周期存储策略,从而保障数据存储策略能够更加符合业务、系统的实际需求,有效地发挥数据生命周期管理的价值,如图4-1所示。

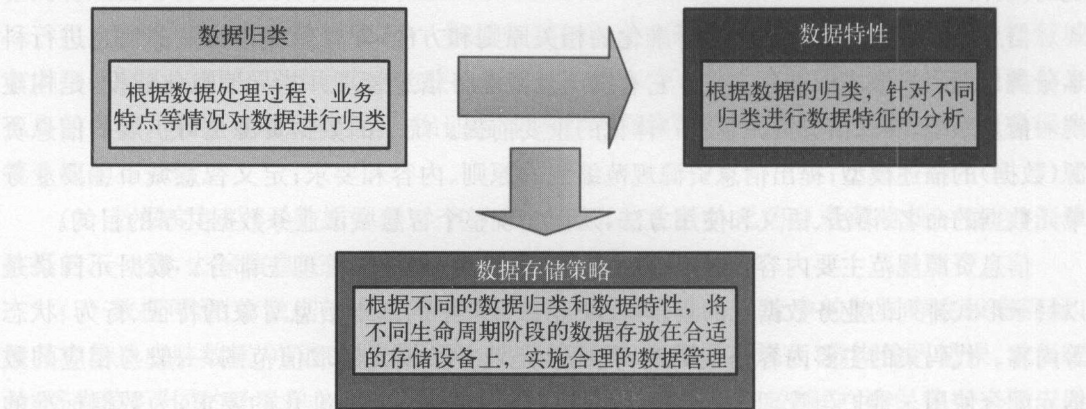


图 4-1 数据存储策略

1) 数据归类

数据归类是从数据生命周期管理视角,结合业界的相关标准规范,对智慧城市各个业务系统的数据进行归类。目前比较流行的数据归类包括:生产交易型数据、服务支撑型数据与系统数据三部分。

(1) 生产交易型数据:就是智慧城市各业务单元系统所产生、使用的数据,该类数据主要分布在业务支撑系统、管理支撑系统。目前生产交易型数据主流的数据归类为:客服域、资源域、企业管理域、营销域、产品域等。

(2) 服务支撑型数据:指从智慧城市各业务单元收集的数据,经沉淀加工挖掘后形成的信息数据,该类数据主要分布在提供数据服务、数据支撑的IT平台中。

(3) 系统数据:智慧城市业务系统在运行过程中所形成的数据,该类数据如果不有效管理将会严重影响系统的运行效率。主要包括:日志数据、过程数据等。

2) 数据特征

根据数据归类,不同类别的数据有不同的数据特性。从数据生命周期视角来看,数据特性主要包括数据重要性、访问频率、访问性能要求、数据量等。具体在进行数据特性分类过程中,可以根据实际情况进行特性归类。

(1) 重要性:按照数据的重要程度进行归类,数据分成不同的重要级别。

(2) 访问频率:依据数据的访问频率,对数据进行归类。通过不同的访问频率定义不同的存储策略。

(3) 访问性能要求:对于不同数据需要提供不同的访问性能支持。通过不同访问性能需要定义不同存储策略。

(4) 数据量:数据以时间单位进行分析,按照数据量的大小,提供不同的存储策略。

3) 数据存储策略

数据存储策略就是将不同的数据存在指定的存储设备上。目前存储策略主要分为在线存储、近线存储和归档存储。

(1) 在线存储:存储设备和所存储的数据时刻保持随时响应状态,可以实时读写,可满足计算平台对数据访问的速度要求。一般在线存储设备通常为内置磁盘和高端磁盘阵列等磁盘设备,价格相对昂贵,性能最好。

(2) 近线存储:定位于在线存储和归档存储之间的应用,性能要求相对来说并不高,但又提供相对较好的读取性能。近线存储策略一般采用中低端磁盘阵列设备,并辅以高压缩软件,满足快速读写等访问动作。

(3) 归档存储:将不再经常使用的数据移到一个单独的存储设备来进行长期保存的过程,对涉及的数据进行离线存储,以备非常规查询等。归档设备一般采用磁带库、光盘库,以及大容量低端磁盘阵列,价格相对低廉。

存储策略的制定除了满足业务需要之外,还需要考虑存储成本,制定合理的存储策略。例如:生产交易型数据中的客服数据,其数据特点主要是数据重要性高、访问频率高、数据

量较小,可以采用在线存储;服务支撑型中的原型数据,其数据特点主要是数据重要性不高、访问频率不高、数据量大,可以采用近线存储。

4.2 大数据采集

在大数据时代,数据无疑是智慧城市建设的核心资产之一,若能盘活好数据,则能使城市治理、城市决策和市民服务等方方面面受益匪浅;反之,则容易导致其核心竞争力下降,甚至衰落。

4.2.1 网络数据获取

网络数据获取是指通过网络爬虫(Web Spider)或网站公开 API 等方式从网站上获取数据信息。该方法可以将非结构化数据从网页中抽取出来,将其存储为统一的本地数据文件,并以结构化的方式存储。它支持图片、音频、视频等文件或附件的采集,附件与正文可以自动关联。

网络数据获取是利用互联网搜索引擎技术实现有针对性、行业性、精准性的数据抓取,按照一定规则和筛选标准进行数据归类,并形成数据库文件的一个过程。网络数据获取是应用大数据的第一步,所谓“巧妇难为无米之炊”,怎样在海量的网络信息中尽可能多地抓取到用户想要的信息,是网络数据获取的核心问题。

网络爬虫是一种按照一定的规则,自动抓取网络信息的程序或者脚本。网络爬虫通过网页的链接地址来寻找网页,从网站的某个页面开始,读取当前网页内容,并找到网页中的其他链接地址(URL),然后通过这些 URL 进入下一个网页,如此循环下去,直到把这个网站所有的网页抓取完成或者遇到停止条件为止。

网络爬虫抓取数据的网站,按照抓取特点可以分为新闻、论坛、微博、视频、黄页、图片、下载类网站。根据不同的网络数据类型与网站结构,采取不同的采集规则。下面介绍论坛、微博和图片视频 3 种典型网站的抓取特点。

1) 论坛的数据获取

论坛网站的特点:话题线索信息分散在多个 Web 页面中,无法获得完整统一的语意单元视图;留言页面可能包含大量与页面主题无关的导航和功能链接以及广告内容;一个页面中可能包含多个留言消息等。

针对这些特点,对论坛数据的获取,采用深度优先策略,按照用户浏览顺序,从话题目录页面的链接打开某个话题的一个消息页面,根据论坛站点的逻辑链接结构,识别并提取出相同话题线索的链接,顺序浏览这些链接的页面内容。由于可能存在多个链接指向同一

页面的情况,可以使用唯一索引办法,防止信息重复入库。

2) 微博的数据获取

申请多个微博账号,通过模拟登录来采集微博信息。每个账号随机关注不同行业中有一定粉丝量和微博量的用户。定期爬取微博信息。记录下每个账号关注的用户名单,并对这些用户的活跃度进行监测,当用户活跃度过低时,可以用活跃度高的用户将其替换。未进行数据抽取的空闲时间段,可让账号进行一些随机动作(如转发、评论),保持一定的活跃度。此外,为防止账号被查封,申请一些备用账号。当某个账号被停用时,由备用账号替代,继续跟踪停用账号所关注的所有用户。

3) 图片视频的数据获取

对于图片和视频类网站,数据抓取中要解决的是内容识别问题。一般是通过链接文本和相关的文件注释来判断文件的内容。例如一个链接文字为“成龙的照片”,链接指向一张图片,那么网络爬虫就知道这张图片的内容是成龙的照片。所以在抓取图片和视频信息时,应先抓取链接文本信息,再通过链接抓取图片和视频。

4.2.2 媒体流获取

移动技术最终目标是无论何时何地,人们都可以得到各种媒体和信息的实时传输交换服务。流媒体是将连续的影像和声音等多媒体信息经过压缩编码处理,在 IP 网上采用流式传输技术传输的时基媒体。由于在客户端缓冲存储,流媒体技术有效地降低了对传输带宽、时延和抖动的要求。它是网络音、视频技术发展 to 一定阶段的产物,涉及数据的采集、压缩、编码、存储、传输以及网络通信等多项技术。流媒体技术应用到移动网络和终端上,称之为移动流媒体技术。流媒体与移动通信结合成为移动通信领域内的一个热点,随着移动通信的普及和发展,移动流媒体服务将成为智慧城市系统一个重要的业务服务。

由于流媒体数据量较大,需要考虑只能对比较重要的少量流媒体提供接入服务。考虑到带宽的限制,媒体流接入服务采用集群的方式,为每个接入源动态确定一个存储服务器,将媒体流使用分布式文件存储落地。媒体就接入服务同时生成媒体的描述信息、摘要信息,保存到关系数据库中,供检索使用。

4.2.3 日志信息获取

智慧城市的应用平台每天会产生大量的日志(一般为流式数据,如 PageView、查询等),日志系统的关键点是从不同的日志源上收集相关日志,存储到一个中央存储系统(可以是 NFS、分布式文件系统等)上,以便于进行集中统计分析处理。系统需要为日志的“分布式收集,统一处理”提供了一个可扩展的、高容错的方案。一般而言,这些系统需要具有以下特征:

(1) 可靠性:当节点出现故障时,日志能够被传送到其他节点上而不会丢失。

- (2) 可扩展性: 当数据量增加时, 可以通过增加节点进行水平扩展。
- (3) 可管理性: 统一管理, 使得系统便于维护, 用户可以查看各个数据源或者数据流执行情况, 且可以对各个数据源配置和动态加载。
- (4) 功能可扩展: 针对不同的日志源, 用户可以自行定义和开发相应的适配器, 扩展整个日志系统的功能度。
- (5) 松耦合性: 构建应用系统和分析系统的桥梁, 并将它们之间的关联解耦。
- (6) 易用性: 支持近实时的在线分析系统和类似于 Hadoop 之类的离线分析系统。

很多互联网企业都有自己的海量数据采集工具, 多用于系统日志采集, 如 Hadoop 的 Chukwa、Cloudera 的 Flume、Facebook 的 Scribe 等, 这些工具均采用分布式架构, 能满足每秒数百兆的日志数据采集和传输需求。

4.2.4 传统应用数据获取

在建设智慧城市过程中, 需要尽可能地使用以前的 IT 遗产, 其中最重要的就是运行了很多年的城市业务系统所积累的海量数据, 而这些数据几乎都是存储于传统的关系型数据库中, 将这部分数据集中到大数据平台的工作就显得非常重要。

和传统方式一样, 大数据的获取过程中, 要将大量的、不同数据库中的信息抽取出来, 放入采集数据库中, 以便分析和挖掘。来自各种不同数据库系统的大量信息和各式各样的分析挖掘需求, 使得对数据交换和处理的要求更高。采用 ETL 将不同数据库中的数据按照统一的模型集成整合到采集数据库中, 是获取数据的重要方式。

ETL 用来描述将数据从来源端经过抽取(Extract)、转置(Transform)、加载(Load)到目的端的过程。

- (1) 抽取。数据抽取是从数据源中提取数据的过程。根据不同的数据源, 采用不同的数据抽取方式。
- (2) 转置。将从不同数据源抽取的不规范数据经过转换和加工, 如去躁、规范格式等, 转换为符合目标数据库的存储格式。
- (3) 加载。将转换和加工后的数据装载到目标数据库中。

ETL 将分布的、异构的数据如关系数据、平面数据文件等抽取到临时中间层后进行清洗、转换、集成, 最终按照定义好的数据模型, 将数据加载到目标数据库中。

4.3 大数据存储

大数据存储层作为智慧城市中超大规模数据信息支撑的基础, 是智慧城市基础设施的

一个重要组成部分。现有的以大规模、高性能与可扩展为主要基础的网络化分布式海量存储系统的设计理念与研究具有局限性,已经很难满足智慧城市中多种复杂应用在规模、效率、可靠性、能耗、安全和智能化等方面的综合性存储服务能力的需求。归纳起来,智慧城市的数据存储需求具有以下几个特点:

(1) 在智慧城市中的一些应用(如政府信息化数据平台、数字交通、环境监测、灾难信息处理等)的数据量以 TB(1 000 GB)级和 PB(1 000 TB)级的速度在高速增长,这种数据规模和数据管理的复杂度已经今非昔比。

(2) 各种应用对数据的可访问时间和访问持续时间在增加,使得数据必须保证能够时刻处于可访问状态。

(3) 数据种类和数据应用的不断繁衍,海量数据呈现动态生长、异构多样化、地理分散、语义丰富等特征。

(4) 数据应用的用户广泛分布,用户访问方式存在多样化与群体众多性。

(5) 应用需要可靠、高效低耗与经济性相结合的数据存储。

智慧城市中数据存储的按需视图特性需要高可扩展性的支持,不仅要为存储本身提供可扩展性(功能扩展),而且必须为存储带宽提供可扩展性(负载扩展)。云存储的另一个关键特性是数据的地理分布(地理可扩展性),支持经由一组云存储数据中心(通过迁移)使数据最接近于用户。对于只读数据,也可以采用多副本和数据布局技术来提高访问性能。海量数据存储策略如图 4-2 所示。

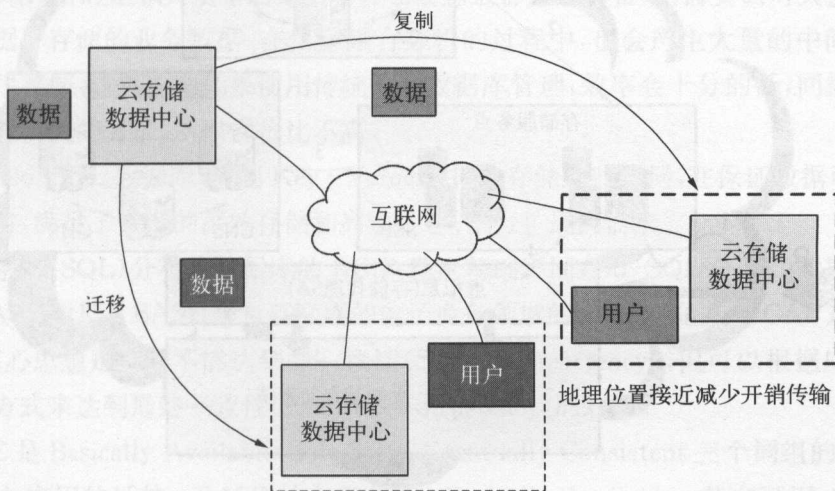


图 4-2 海量数据存储策略

智慧城市存储层的技术体系架构如图 4-3 所示,通过虚拟化存储技术可以将分布在互联网上的各种存储资源整合成具有统一逻辑视图的高性能大规模存储系统。图中的 SSP (Storage Service Point, 存储服务点)是整个存储层的入口,对存储层所有模块的访问都通过 SSP,它主要提供 FTP (File Transfer Protocol, 文件传输协议) 接口、CA (Certificate

Authority, 认证授权) 接口、RM (Resource Manager, 资源管理器) 接口和 GNS (Global Name Server, 全局命名服务器) 接口。在整个存储层中 SSP 不是唯一的, 系统可以根据需要动态增加, SSP 实现了传统存储系统中命名服务器的部分功能, 减轻了 GNS 的负载, 提高了存储层的可扩展性。GNS 负责存储层的元数据管理, 主要包括元数据操作接口、元数据容错系统、元数据搜索系统。RM 包括资源调度模块和副本管理模块, 主要负责资源的申请和调度, 同时提供透明的副本创建和选择策略。副本技术减少了文件数据访问延迟和带宽消耗, 有助于改善负载平衡和可靠性。尤其是动态的副本创建机制, 即自动的选择存储点以创建副本, 并根据用户的特征而自动变化创建策略, 为副本机制提供了更高的灵活性。客户端将支持 3 种形式: 通用 FTP 客户端、文件访问接口和特制客户端。用户通过系统提供的特制客户端, 不但能够进行用户组操作, 具有搜索和共享等功能, 还可以获得更高性能的服务。CA 包含证书管理系统, 主要负责存储层的安全性和数据的访问控制, 同时它记录了用户的注册信息。虚拟层的存储代理 SA 屏蔽了存储资源和设备的多样性, 为系统提供统一存储访问接口, 同时提供了文件操作方式和扩展的 FTP 操作方式, 另外它对文件复制管理操作提供支持, 为高效传输提供服务。同时 SA 这一级实现了局域存储资源的虚拟化, 包括统一 SAN (存储区域网) 和 NAS (网络附加存储)、分布式的磁盘虚拟化、磁带库虚拟化和 SAN 内部共享管理等。

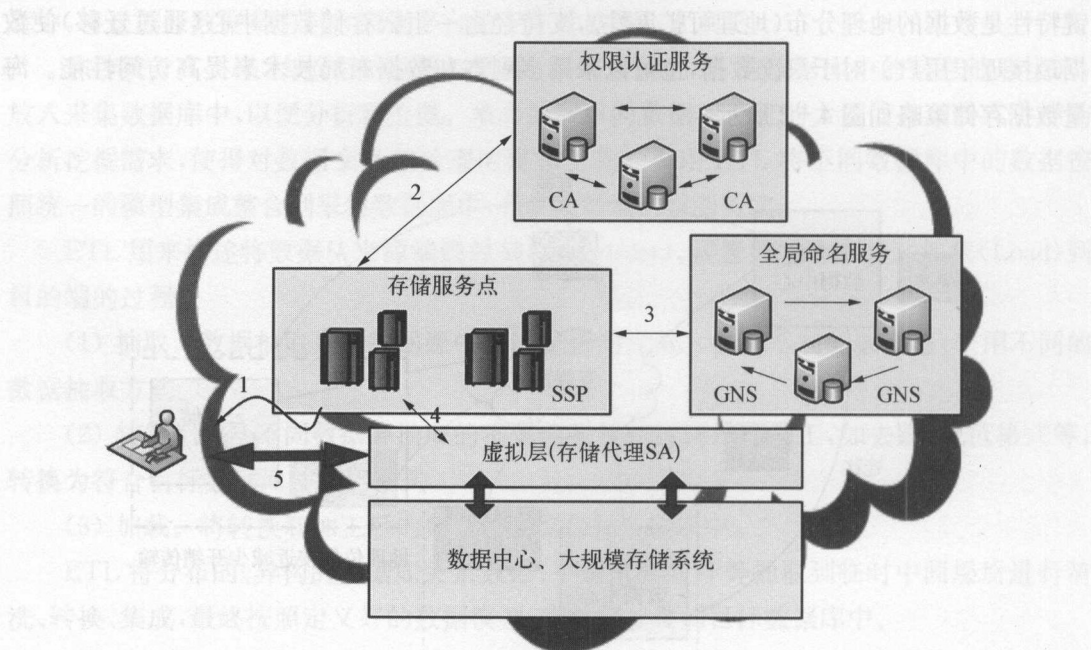


图 4-3 智慧城市存储层的技术体系架构

4.3.1 存储技术

智慧城市大数据分析中的处理对象既包含结构化数据, 也包括非结构化、半结构化数

据,所以从存储技术的角度,关系型数据库、No-SQL、分布式文件系统都应纳入我们的考虑范畴。下面就这些主流存储技术在大数据应用中的定位和使用一一解释。

1) 关系型数据库

智慧城市建设包括大量传统的信息化系统的建设,关系数据库是传统信息化系统的数据基础。通过异构数据交换平台,从各业务系统中获取数据并存储。

关系型(SQL)分布式数据库的理论基础是 ACID 模型,即原子性(Atomicity)、一致性(Consistency)、隔离性(Isolation)和持久性(Durability)的缩写。关系型数据库的理论基础,决定了其具有较强的并发读写能力、数据强一致性保障、很强的结构化查询与复杂分析能力和标准的数据访问接口,非常适合企业私有云数据库平台的建设,是处理结构化数据的主流技术。

然而随着互联网应用的高速发展,数据库的大小与以往局域网内的应用相比提高了若干个数量级。通过 SQL 数据库的 ACID 可以知道传统的关系型数据库因为通用性设计带来了性能上的限制,常见的解决方法主要由以下两种:

(1) 通过集群提供较强的横向扩展能力,但是当节点增加一定数量时,性能上就难以获得明显的提升了。而且这种集群的成本很高。

(2) 数据分片技术,通俗地讲,分片是将一个大数据库按照一定规则拆分成多个小数据库的一门技术。常用的分片方案有以下两种:按功能划分(垂直切分)、按表中某一字段值的范围划分(水平切分)。

2) No-SQL 数据库

No-SQL 指的是非关系型的数据库。在信息融合处理平台中,需要面对大量不适合传统关系数据库存储的业务数据,在信息融合分析的过程中,也会产生大量的中间数据需要高效的顺序存储,这些数据如果使用传统关系数据库管理,效率会十分的低,同样不能满足数据量平行扩展的需求,所以性价比不高。

通过 No-SQL 数据库,使用 Key-Value 结构来存储这些数据,在保证数据可用性和关系型的同时,满足了数据的高效存储和高速处理、平行扩展的需求。

和关系型(SQL)分布式数据库的 ACID 理论基础不同,No-SQL 分布式数据库的理论基础是 BASE 模型。BASE 来自互联网的电子商务领域的实践,它是基于 CAP 理论逐步演化而来,核心思想是即便不能达到强一致性(Strong Consistency),但可以根据应用特点采用适当的方式来达到最终一致性(Eventual Consistency)的效果。

BASE 是 Basically Available、Soft State、Eventually Consistent 三个词组的简写,是对 CAP 中 CA 应用的延伸。BASE 的含义包括:Basically Available:基本可用;Soft State:软状态/柔性事务,即状态可以有一段时间的不同步;Eventually Consistency:最终一致性;BASE 是反 ACID 的,它完全不同于 ACID 模型,牺牲强一致性,获得基本可用性和柔性可靠性并要求达到最终一致性^[1]。

3) 实时数据库

在信息融合处理平台中,主要应用场景中所面对的数据实时性要求无须达到工业现场

数据采集的级别,平台所提供的实时性通常是在 5 s 左右的级别上对采集到的实时数据进行处理分析,并提供给业务系统使用。针对这一需求,从结构上参考了传统实时数据库的实现方式,并根据智慧城市信息融合处理平台的特殊业务需求,将实时数据的存储与实时数据的获取分离实现。在实时数据的存储上,将失效的历史数据与有效的当前数据分别存储。

内存数据库是支持实时任务的最佳技术。内存数据库以内存作为数据库的存储介质,因而存取速度快,和 CPU 之间数据交换快速,避免了数据库存储在外存的速度慢和执行时间不可预知等缺点。内存数据库要求有较大内存,但并不要求任何时刻的数据和所有数据都放入内存。具有短暂有效时间的数据必须保存在内存、存取频率高的数据保存在内存。内存数据库要具有 I/O 接口,同存储在外部存储器的关系数据库交换数据。选用内存数据库来支持实时事务的实现,它将数据驻留在内存当中,消除了传统的磁盘数据库事物运行的 I/O 瓶颈,从而获得 CPU 直接访问数据库数据的极高存取速度,极大地提高了系统性能和吞吐量。

针对有效的当前数据,应用需要尽可能快地提供数据的索引和存储,以提供分析使用,所以采用内存数据库的方式完成对有效数据的存储,数据的保存采用顺序保存的方式。内存数据库的时效引擎会自动扫描有效数据,失效的数据会被顺序转入历史数据库,使用 Key - Value 结构保存。

对于历史数据,平台提供了多种不同的压缩算法,以适应业务系统的需要。包括有损的线性拟合压缩算法、无损的哈夫曼编码方式的压缩算法,可节约大量的存储空间。

4) 分布式文件存储

相对于结构化数据而言,不方便用数据库二维逻辑表来表现的数据即称为非结构化数据,包括所有格式的办公文档、文本、图片、XML、HTML、各类报表、图像和音频/视频信息,等等。据 IDC 的研究人员表示,如今 80% 的企业数据是非结构化数据,并且以 60% 的年增长率在增加。非结构化数据是把信息存储在文件系统中,而不是数据库中。重要的是,研究人员称,经常性使用的数据比例平均只有 1%~5%。

分布式文件系统(Distributed File System)是实现非结构化数据存储的主要技术。分布式文件系统是指文件系统管理的物理存储资源不一定直接连接在本地节点上,而是通过计算机网络与节点相连。分布式文件系统的设计基于客户机/服务器模式。一个典型的网络可能包括多个供多用户访问的服务器。另外,对等特性允许一些系统扮演客户机和服务器的双重角色。例如,用户可以“发表”一个允许其他客户机访问的目录,一旦被访问,这个目录对客户机来说就像使用本地驱动器一样。

目前分布式文件系统的实施标准和实现是 Apache Hadoop 的 HDFS 系统,HDFS 也是基于谷歌著名的分布式文件系统论文进行构造。它有着高容错性的特点,并且设计用来部署在低廉的硬件上,而且它提供高吞吐量来访问应用程序的数据,适合那些有着超大数据集的应用程序。HDFS 放宽了 POSIX 的要求,这样可以实现流的形式访问文件系统中的

数据。

(1) 超大文件：在这里是指具有几百兆、几百吉甚至几百太大小的文件，目前已经有存储几百帕级别数据的 Hadoop 集群了。

(2) 流式数据访问：HDFS 的思路是一次写入、多次读取，是最为高效的访问模式，数据集通常由数据源生成或从数据源复制而来，接着长时间在此数据集上进行各类分析。每次分析都将设计到该数据集的大部分甚至全部数据。

(3) 商用硬件：Hadoop 并不需要运行在昂贵并且高可靠的硬件上，它是被设计运行在商用硬件的集群上的。

(4) 低时间延迟的数据访问：要求低时间延迟数据访问的应用，例如几十毫秒范围，不适合在 HDFS 上运行，HDFS 是为高数据吞吐量应用优化的，这可能会以高时间延迟为代价。

5) 列式数据库

列式数据库是以列相关存储架构进行数据存储的数据库，主要适合于批量数据处理和即席查询。面向列的数据存储架构更适用于 OLAP 这样在海量数据(可能达到太字节规模)中进行有限复杂查询的场景。列式数据库对比传统的关系型行式数据库的优势主要体现在以下两个方面：

(1) 不读取无效数据：降低 I/O 开销，同时提高每次 I/O 的效率，从而大大提高查询性能。查询语句只从磁盘上读取所需要的列，其他列的数据是不需要读取的。

(2) 高压缩比：压缩比可以达到 5~20 倍，数据占有空间降低到传统数据库的十分之一，节省了存储设备的开销。当数据库的大小与数据库服务器内存大小之比达到或超过 2:1(典型的大型系统配置值)时，列存的 I/O 优势就显得更加明显。

目前主流的列式数据库方案是 Apache Hadoop 的 HBase 体系，和 HDFS 一样，HBase 也是基于谷歌的 BigTable 论文进行构造的。

4.3.2 硬件架构

1) 本地存储

本地存储就是本地磁盘，是指安装于同一台计算机主板上，不可随意插拔、移动的磁盘(硬盘)，一般包括计算机操作系统所在分区及其他分区。本地磁盘区别于通过网络连线访问的共享磁盘(如服务器磁盘或网络上其他计算机的共享磁盘)或通过本计算机外部连线连接但可随意插拔的移动磁盘(硬盘)。同一台计算机内部安装的多块磁盘均可称作该计算机的本地磁盘。

在大多数情况下，基于服务器的本地存储并没有高效率的优点。但本地存储在两个方面占据优势：成本和整体性能。问题是当服务器依赖于本地存储时，如何在必要的时候增加更多的容量。通常，有两种方式来处理这种困境。第一种方法是增加具有更多本地存储

的额外的服务器。第二种方法是增加集群服务器的容量。这两种方法都需要购买和配置硬件,这将导致停机时间,可能还需要重新设计架构。

然而,基于服务器的磁盘也存在缺点,主要是在可扩展性方面。

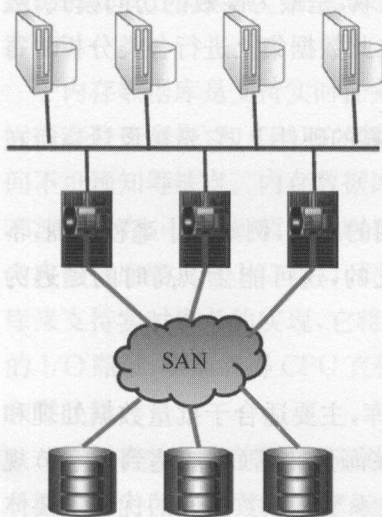


图 4-4 SAN 共享存储架构

2) SAN 存储

使用高速光纤结构的 SAN 是共享存储架构,是一种高速网络或子网络,提供在计算机与存储系统之间的数据传输。存储设备是指一张或多张用以存储计算机数据的磁盘设备。一个 SAN 网络由负责网络连接的通信结构、负责组织连接的管理层、存储部件以及计算机系统构成,从而保证数据传输的安全性和力度,如图 4-4 所示。

SAN 解决方案通常会采取以下两种形式:光纤信道以及 iSCSI 或者基于 IP 的 SAN。光纤信道是 SAN 解决方案中大家最熟悉的类型,但是,最近一段时间以来,基于 iSCSI 的 SAN 解决方案开始大量出现在市场上,与光纤通道技术相比较而言,这种技术具有良好的性能,而且价格低廉。

SAN 真正地综合了 DAS(Direct-Attached Storage, 开放系统的直连式存储)和 NAS 两种存储解决方案的优势。例如,在一个很好的 SAN 解决方案实现中,可以得到一个完全冗余的存储网络,这个存储网络具有不同寻常的扩展性,确切地说,你可以得到只有 NAS 存储解决方案才能得到的几百太字节的存储空间,但是你还可以得到块级数据访问功能,而这些功能只能在 DAS 解决方案中才能得到。对于数据访问来说,你还可以得到一个合理的速度,对于那些要求大量磁盘访问的操作来说,SAN 具有更好的性能。利用 SAN 解决方案,你还可以实现存储的集中管理,从而能够充分利用那些处于空闲状态的空间。更有优势的一点是,你甚至可以将服务器配置为没有内部存储空间的服务,要求所有的系统都直接从 SAN(只能在光纤通道模式下实现)引导。这也是一种即插即用技术。

SAN 的缺陷在于其成本、复杂性及规模有限。而且,SAN 方法本身并不一定适合大数据部署。尤其是那些使用 Hadoop 的大数据部署,因为 SAN 承担集中硬盘上数据的责任,这反过来意味着每个计算服务器将需要访问相同的 SAN 来恢复正态分布的数据。

3) 云存储

云存储的概念与云计算类似,它是指通过集群应用、网格技术或分布式文件系统等功能,将网络中大量各种不同类型的存储设备通过应用软件集合起来协同工作,共同对外提供数据存储和业务访问功能的一个系统。

云存储系统的结构模型由 4 层组成。

(1) 存储层:云存储最基础的部分。存储设备可以是 FC 光纤通道存储设备,可以是

NAS 和 iSCSI 等 IP 存储设备,也可以是 SCSI 或 SAS 等 DAS 存储设备。云存储中的存储设备往往数量庞大且分布于不同地域,彼此之间通过广域网、互联网或者 FC 光纤通道网络连接在一起。存储设备之上是一个统一存储设备管理系统,可以实现存储设备的逻辑虚拟化管理、多链路冗余管理,以及硬件设备的状态监控和故障维护。

(2) 基础管理层:云存储最核心的部分,也是云存储中最难以实现的部分。基础管理层通过集群、分布式文件系统和网格计算等技术,实现云存储中多个存储设备之间的协同工作,使多个存储设备可以对外提供同一种服务,并提供更大、更强、更好的数据访问性能。CDN 内容分发系统、数据加密技术保证云存储中的数据不会被未经授权的用户所访问,同时,通过各种数据备份以及容灾技术和措施,可以保证云存储中的数据不会丢失,保证云存储自身的安全和稳定。

(3) 应用接口层:云存储最灵活多变的部分。不同的云存储运营单位可以根据实际业务类型,开发不同的应用服务接口,提供不同的应用服务。比如视频监控应用平台、IPTV 和视频点播应用平台、网络硬盘引用平台、远程数据备份应用平台等。

(4) 访问层:任何一个授权用户都可以通过标准的公用应用接口来登录云存储系统,享受云存储服务。云存储运营单位不同,云存储提供的访问类型和访问手段也不同。

4.3.3 CAP 原则及技术选择

在存储技术的不同方案比较中,已经提到过 CAP 理论,下面详细阐述 CAP 原则并基于该原则对现有大数据存储方案进行比较和考量。

在理论计算机科学中,CAP 定理又被称作布鲁尔定理(Brewer's Theorem),它指出对于一个分布式计算系统来说,不可能同时满足以下 3 点:

- (1) 一致性(Consistency):所有节点在同一时间具有相同的数据。
- (2) 可用性(Availability):保证每个请求不管成功或者失败都有响应。
- (3) 分区容忍(Partition Tolerance):系统中任意信息的丢失或失败不会影响系统的继续运作。

在讨论 CAP 的时候,最为关心的是一致性模型的问题。数据的一致性,一般来说有 3 种类型:弱一致性、最终一致性和强一致性。

(1) 弱一致性(Weak):当你写入一个新值后,读操作在数据副本上可能读出来,也可能读不出来。比如:某些 Cache 系统、VOIP 系统等。

(2) 最终一致性(Eventually):当你写入一个新值后,有可能读不出来,但在某个时间窗口之后保证最终能读出来。比如:DNS、电子邮件系统等。

(3) 强一致性(Strong):新的数据一旦写入,在任意副本任意时刻都能读到新值。比如:文件系统、关系型数据库都是强一致性的。

从这三种一致型模型可以看到,弱一致性和最终一致性一般来说是异步冗余的,而强

一致性一般来说是同步冗余的。需要注意的是,异步通常意味着更好的性能,但也意味着更复杂的状态控制。同步意味着简单,但也意味着性能下降。

1) Master-Slave(主机-备机)架构

首先是 Master-Slave 结构,对于这种架构,Slave 一般是 Master 的备份。在这样的系统中,一般是如下设计的:

(1) 读写请求都由 Master 负责。

(2) 写请求写到 Master 上后,由 Master 同步到 Slave 上。

从 Master 同步到 Slave 上,可以使用异步,也可以使用同步,可以使用 Master 来 push(推),也可以使用 Slave 来 pull(拉)。通常来说是 Slave 来周期性的 pull,所以,是最终一致性。这个设计的问题是,如果 Master 在 pull 周期内垮掉了,那么会导致这个时间片内的数据丢失。如果不想让数据丢掉,Slave 只能成为 Read-Only 的方式等 Master 恢复。

当然,如果可以容忍数据丢掉的话,可以马上让 Slave 代替 Master 工作(对于只负责计算的节点来说,没有数据一致性和数据丢失的问题,Master-Slave 的方式就可以解决单点问题了)。当然,Master Slave 也可以是强一致性的,比如:写 Master 的时候,Master 负责先写自己,成功后,再写 Slave,两者都成功后返回成功,整个过程是同步的,如果写 Slave 失败了,那么两种方法,一种是标记 Slave 不可用报错并继续服务(等 Slave 恢复后同步 Master 的数据,可以有多个 Slave,这样少一个,还有备份,就像前面说的写三份那样),另一种是回滚自己并返回写失败。可以看到,如果 Master-Slave 需要做成强一致性有多复杂。

2) Master-Master 架构

Master-Master,又叫 Multi-Master,是指一个系统存在两个或多个 Master,每个 Master 都提供 Read-Write(读-写)服务。这个模型是 Master-Slave 模型的增强,数据间同步一般是通过 Master 间的异步完成,所以是最终一致性。Master-Master 的好处是,一台 Master 宕机了,别的 Master 还是可以正常做读写服务,它和 Master-Slave 一样,当数据没有被复制到别的 Master 上时,数据会丢失。很多数据库都支持 Master-Master 的 Replication(复制)机制。

另外,如果多个 Master 对同一个数据进行修改的时候,这个模型的噩梦就出现了——对数据间的冲突合并,这并不是一件容易的事情。看看 Dynamo 的 Vector Clock 的设计(记录数据的版本号和修改者)就知道这个事并不那么简单,而且 Dynamo 对数据冲突这个事是交给用户自己处理的。就像我们的 SVN 源码冲突一样,对于同一行代码的冲突,只能交给开发者自己来处理。

3) 两段式提交

在分布式系统中,每个节点虽然可以知晓自己的操作是成功或者失败,却无法知道其他节点操作的成功或失败。当一个事务跨越多个节点时,为了保持事务的 ACID 特性,需要

引入一个作为协调者的组件来统一掌控所有节点(称作参与者)的操作结果并最终指示这些节点是否要把操作结果进行真正的提交(比如将更新后的数据写入磁盘等)。两阶段提交的算法如下:

(1) 第一阶段:

① 协调者会问所有的参与者节点,是否可以执行提交操作。

② 各个参与者开始事务执行的准备工作,如为资源上锁,预留资源,写 undo/redo log……

③ 参与者响应协调者,如果事务的准备工作成功,则回应“可以提交”,否则回应“拒绝提交”。

(2) 第二阶段:

① 如果所有的参与者都回应“可以提交”,那么,协调者向所有的参与者发送“正式提交”的命令。参与者完成正式提交,并释放所有资源,然后回应“完成”,协调者收集各节点的“完成”回应后结束这个 Global Transaction。

② 如果有一个参与者回应“拒绝提交”,那么,协调者向所有的参与者发送“回滚操作”,并释放所有资源,然后回应“回滚完成”,协调者收集各节点的“回滚”回应后,取消这个 Global Transaction。

4) PAXOS 算法

PAXOS 算法解决的问题是在一个可能发生上述异常的分布式系统中,如何就某个值达成一致,保证不论发生以上任何异常,都不会破坏决议的一致性。一个典型的场景是,在一个分布式数据库系统中,如果各节点的初始状态一致,每个节点都执行相同的操作序列,那么他们最后能得到一个一致的状态。为保证每个节点执行相同的命令序列,需要在每一条指令上执行一个“一致性算法”,以保证每个节点看到的指令一致。一个通用的一致性算法可以应用在许多场景中,是分布式计算中的重要问题。

简单说来,PAXOS 的目的是让整个集群的节点对某个值的变更达成一致。PAXOS 算法基本上来说是个民主选举的算法——大多数的决定会成整个集群的统一决定。任何一个点都可以提出要修改某个数据的提案,是否通过这个提案取决于这个集群中是否有超过半数的节点同意(所以 PAXOS 算法需要集群中的节点是单数)。

PAXOS 算法可以简单地看作是对两段式提交的优化和完善,当然自从 Lamport 在 1998 年发表 PAXOS 算法后,对 PAXOS 的各种改进完善工作也从未停止过,其中动作最大的莫过于 2005 年发表的 Fast PAXOS。无论何种改进,其重点依然是在消息延迟与性能、吞吐量之间作出各种权衡。

5) CAP 总结和技术选型

综上所述,要想让数据有高可用性,就需要冗余数据写多份。写多份的问题会带来一致性的问题,而一致性的问题又会带来性能问题。从图 4-5 可以看到,基本上不可以让所有的项都绿起来,这就是著名的 CAP 理论:一致性、可用性、分区容忍性,你只可能要其中

	备份	M-S	M-M	两阶段提交	PAXOS
一致性	弱	最终一致性		强	
事务	无	全部支持	本地事务	全部支持	
等待时间	低			高	
吞吐量	高			低	中等
数据丢失	大量	一些		无	
故障容错	无	只读	读/写		

图 4-5 CAP 总结和技术选型

的两个。

(1) 一致性,系统中对一个数据的读和写虽然包含多个子步骤并且会持续一段时间才能执行完,但是在调用者看来,读操作和写操作都必须是单个的即时完成的操作,不存在重叠。对一个写操作,如果系统返回了成功,那么之后到达的读请求都必须读到这个新的数据;如果系统返回失败,那么所有的读,无论是之后发起的,还是和写同时发起的,都不能读到这个数据。

(2) 可用性和分区容错性的前提是必须要定义好系统的故障模型。在形式化证明中,系统包含多个节点,每个节点可以接收读和写的请求,返回成功或失败,对读还要返回一个数据,和调用者之间的连接是不会中断的,系统的节点也不会失效,唯一的故障就是报文的丢失。

目前主流的大数据存储方案以及 CAP 特性分析如图 4-6 所示。

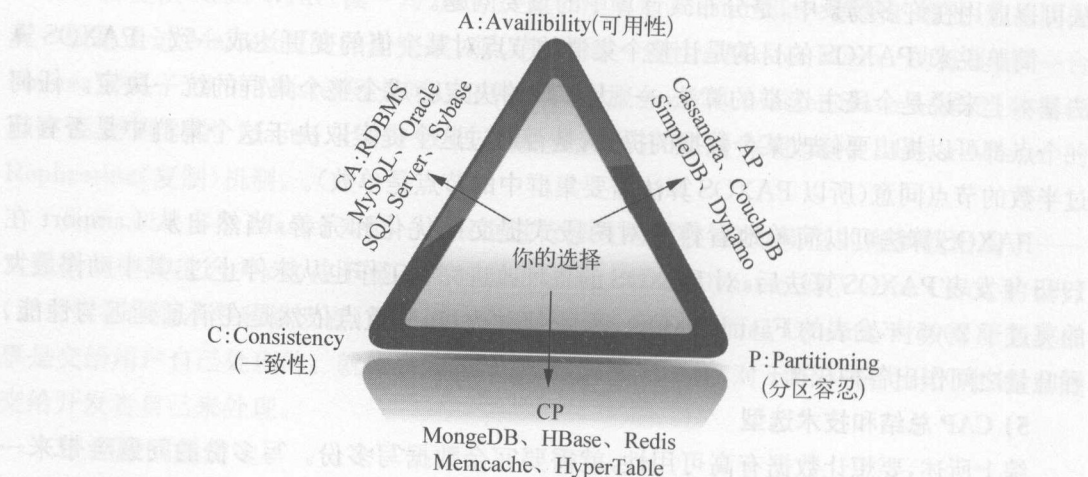


图 4-6 CAP 特性和相应存储方案选择

4.3.4 大数据存储的未来

1) Google Spanner 体系：冲破 CAP

Spanner 是谷歌的全球级的分布式数据库 (Globally-Distributed Database)。Spanner 的扩展性达到了令人咋舌的全球级, 可以扩展到数百万的机器、数以百计的数据中心、上万亿的行。更进一步, 它还能同时通过同步复制和多版本来满足外部一致性, 可用性也是很好的。谷歌尝试冲破 CAP 的枷锁, 在三者之间做到完美平衡。

Spanner 是谷歌的第一个可以全球扩展并且支持外部一致的事务。能做到这些, 得益于一个用 GPS 和原子钟实现的时间 API。这个 API 能将数据中心之间的时间同步精确到 10 ms 以内, 还具有几个重要的功能: 无锁读事务、原子数据结构修改、无阻碍地读取历史数据。

从高层看 Spanner 是通过 PAXOS 状态机将分区好的数据分布在全球的, 用户可以指定数据复制的份数和存储的地点。Spanner 可以在集群或者数据发生变化时将数据迁移到合适的地点, 做负载均衡。用户可以指定将数据分布在多个数据中心, 不过更多的数据中心将造成更多的延迟。用户需要在可靠性和延迟之间做权衡, 一般来说复制 1~2 个数据中心足以保证可靠性。Spanner 由于是全球化的, 所以有两个其他分布式数据库没有的概念。

(1) 宇宙 (Universe): 一个 Spanner 部署实例称为一个 Universe, 目前全球有三个。一个开发, 一个测试, 一个线上。一个 Universe 就能覆盖全球, 不需要多个。

(2) 区段 (Zones): 每个 Zone 相当于一个数据中心, 一个 Zone 内部物理上必须在一起。而一个数据中心可能有多个 Zone。可以在运行时添加移除 Zone。一个 Zone 可以理解为一个 BigTable 部署实例。

作为一个全球化分布式系统, Spanner 提供一些有趣的特性。

(1) 应用可以细粒度地指定数据分布的位置。精确地指定数据离用户有多远, 可以有效地控制读延迟 (读延迟取决于最近的拷贝)。指定数据拷贝之间有多远, 可以控制写的延迟 (写延迟取决于最远的拷贝)。还有指定数据的复制份数, 可以控制数据的可靠性和读性能 (多写几份, 可以抵御更大的事故)。

(2) Spanner 还有两个一般分布式数据库不具备的特性: 读写的外部一致性, 基于时间戳的全局的读一致。这两个特性可以让 Spanner 支持一致的备份, 一致 Map-Reduce, 还有原子的数据结构修改。

2) Amazon NWR 模型：给客户选择 CAP

Amazon Dynamo 的 NWR 模型把 CAP 的选择权交给了用户, 让用户自己的选择你的 CAP 中的哪两个。

作为一个分布式数据存储, 一般工业界认为比较安全的备份数应该是 3 份。而在备份

控制的时候,会考虑到一些问题。

(1) 首先,如何选择备份节点。可以简单地选择顺序上的后两个节点为备份节点,比如存在节点 A 的数据,备份到节点 B 和 C。但是当前面引入了虚节点的概念的时候,有可能 C 节点和 A 节点在同一台物理机器上,这个时候就不能选择 C 作为 A 的备份了。

(2) 其次,当一个节点离开系统的时候,比如宕机,这个节点上存储的信息需要继续备份到其他节点上。

(3) 同样,当一个节点加入系统,从其他节点偷了数据后,其他节点也需要相应减少备份数。而一个节点如果只是暂时性的不可达,也就是失效一个很短的时间,那么需要其他节点暂时接管这个节点的工作,在其可用的时候,把数据增量传送回该节点。

在设计上述需求的解决方案的时候,还要考虑各个节点间数据备份是同步还是异步的问题。假设系统要求写请求总是尽可能的成功,那么策略是写任何一个节点成功就认为成功。节点之间的数据通过异步形式达成一致,这个时候读请求可能读不到最新写进去的信息。这个时候写请求可能要耗较多的时间,甚至根本不能完成(如果有节点不可达)。也就是说,系统的一致性、可靠性、原子性、隔离性的问题(ACID)是无法同时达到的,只能在其中作出取舍。

Amazon 的处理方式是把这个选择权交给用户,这就是它的 NWR 模型。所谓 NWR 模型,N 代表 N 个备份,W 代表要写入至少 W 份才认为成功,R 表示至少读取 R 个备份。配置的时候要求 $W+R > N$,就是读取的份数一定要比总备份数减去确保写成功的倍数的差值要大。

NWR 模型的一些设置会造成脏数据的问题,因为这很明显不是像 PAXOS 一样是强一致的,所以,可能每次的读写操作都不在同一个节点上,于是会出现一些节点上的数据并不是最新版本,但却进行了最新的操作。

Amazon Dynamo 引入了数据版本的设计(利用 Vector Clock 算法)。也就是说,如果你读出来数据的版本是 v1,当你计算完成后要回填数据后,却发现数据的版本号已经被别人更新成了 v2,那么服务器就会拒绝你。

4.4 大数据处理

4.4.1 计算技术

1) 轻量级弹性计算平台

随着大数据平台的发展,基于数据密集型应用的计算框架不断出现,从支持离线处理的 Map-Reduce,到支持在线处理的 Storm,从迭代式计算框架 Spark 到六十分处理框架

S4,各个计算框架诞生于不同的公司(或大学实验室),各有所长,各自解决了某一类应用问题。而在大数据的应用情况下,这些框架可能同时被采用。考虑到资源利用率、运维成本、数据共享等因素,公司一般希望将所有这些框架都部署到一个公共集群中,让它们共享集群的资源,并对资源进行统一的使用,同时采用某种资源隔离方案对各个任务进行隔离,这样便诞生了轻量级弹性计算平台,Apache YARN 就是其中最主要的代表,如图 4-7 所示。

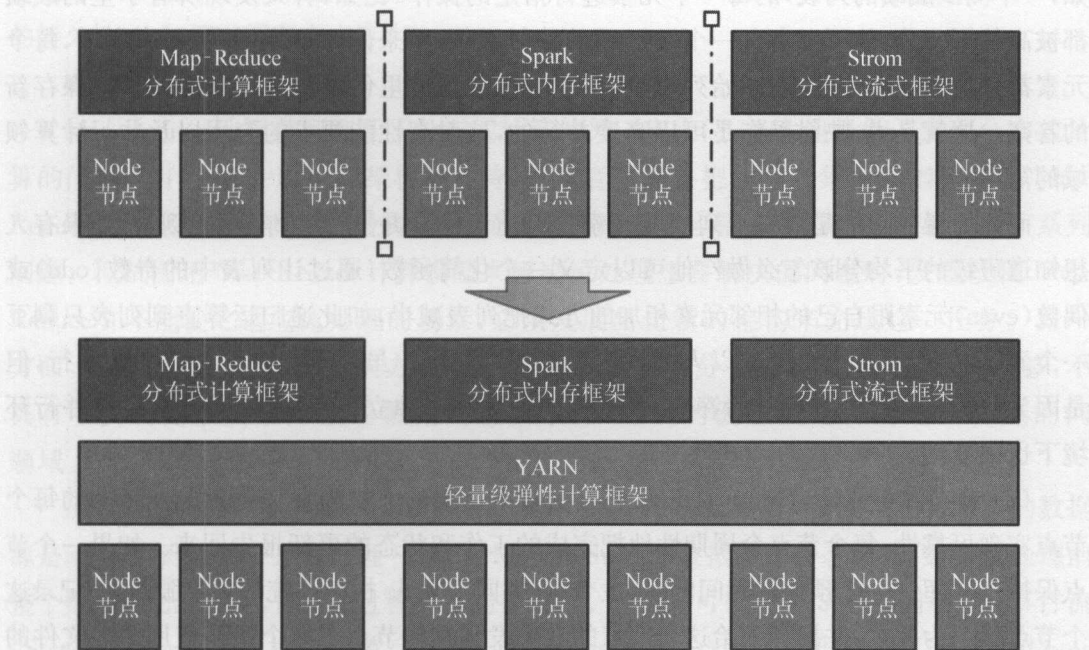


图 4-7 轻量级弹性计算平台

弹性计算平台的引入,它的目标已经不再局限于支持 Map-Reduce 一种计算框架,而是朝着对多种框架进行统一管理的方向发展。相比于“一种计算框架一个集群的模式”,共享集群的模式存在多种好处:

(1) 资源利用率高: 共享集群模式通过多种计算框架共享资源,使得集群中的资源得到更加充分的利用。

(2) 运维成本低: 共享模式通常需要少数管理员即可完成多个框架的统一管理。

(3) 数据共享: 随着数据量的暴增,跨集群间的数据移动不仅需花费更长的时间,并且硬件成本也会大大增加,而共享集群模式可让多种框架共享数据和硬件资源,将大大减少数据移动带来的成本。

2) Map-Reduce 离线计算框架

Map-Reduce 是一种离线计算框架,用于大规模数据集(大于 1 TB)的并行运算。谷歌提出的 Map-Reduce 模型已经成为事实上的并行计算标准。Map-Reduce 主要用于大规模

数据集的并行运算。Map(映射)和 Reduce(化简)反映了并行计算的两个主要的处理环节,对应为两个具体的函数。Map 函数用来把一组键值对映射成一组新的键值对,后续将分发到分布式的处理节点中进行处理,Reduce 函数则用来保证所有映射的键值对中的每一个共享相同的键组,实现多节点处理结果的整合。当然,除了 Map 和 Reduce 外,还包括了 Split、Partition 等过程,这些过程的控制是保证 Map-Reduce 高效执行的关键。

(1) 映射和化简。简单来说,一个映射函数就是对一些独立元素组成的概念列表(例如,一个测试成绩的列表)的每一个元素进行指定的操作(比如,有人发现所有学生的成绩都被高估了一分,他可以定义一个“减一”的映射函数,用来修正这个错误)。事实上,每个元素都是被独立操作的,而原始列表没有被更改,因为这里创建了一个新的列表来保存新的答案。这就是说,映射操作是可以高度并行的,这对高性能要求的应用以及并行计算领域的需求非常有用。

而化简操作指的是对一个列表的元素进行适当的合并[继续看前面的例子,如果有人想知道班级的平均分该怎么做?他可以定义一个化简函数,通过让列表中的奇数(odd)或偶数(even)元素跟自己的相邻元素相加的方式把列表减半,如此递归运算直到列表只剩下一个元素,然后用这个元素除以人数,就得到了平均分]。虽然不如映射函数那么并行,但是因为化简总是有一个简单的答案,大规模的运算相对独立,所以化简函数在高度并行环境下也很有用。

(2) 分布和可靠性。Map-Reduce 通过把对数据集的大规模操作分发给网络上的每个节点实现可靠性;每个节点会周期性地把完成的工作和状态的更新报告回来。如果一个节点保持沉默超过一个预设的时间间隔,主节点(类同 Google 档案系统中的主服务器)记录这个节点状态为死亡,并把分配给这个节点的数据发到别的节点。每个操作使用命名文件的不可分割操作以确保不会发生并行线程间的冲突;当文件被改名的时候,系统可能会把它们复制到任务名以外的另一个名字上去。

化简操作工作方式很类似,但是由于化简操作并行能力较差,主节点会尽量把化简操作调度在一个节点上,或者离需要操作的数据尽可能近的节点上了;这个特性可以满足谷歌的需求,因为他们有足够的带宽,他们的内部网络没有那么多的机器。

3) Spark 内存计算框架

Spark 是一种内存计算集群环境,启用了内存分布数据集,提供了交互式查询。Spark 是在 Scala 语言中实现的,它将 Scala 用作其应用程序框架,而 Scala 的语言特点也铸就了大部分 Spark 的成功。与 Hadoop 不同,Spark 和 Scala 能够紧密集成,其中的 Scala 可以像操作本地集合对象一样轻松地操作分布式数据集。尽管创建 Spark 是为了支持分布式数据集上的迭代作业,但是实际上它是对 Hadoop 的补充,可以在 Hadoop 文件系统中并行运行。通过类似于 YARN 的第三方集群框架支持此行为。

Spark 是为集群计算中的特定类型的工作负载而设计,即那些在并行操作之间重用工作数据集(比如机器学习算法)的工作负载。为了优化这些类型的工作负载,Spark 引

进了内存集群计算的概念,可在内存集群计算中将数据集缓存在内存中,以缩短访问延迟。

Spark 引进了名为弹性分布式数据集(RDD)的抽象。RDD 是分布在一组节点中的只读对象集合。这些集合是弹性的,如果数据集一部分丢失,则可以对它们进行重建。重建部分数据集的过程依赖于容错机制,该机制可以维护“血统”(即允许基于数据衍生过程重建部分数据集的信息)。RDD 被表示为一个 Scala 对象,并且可以从文件中创建它;一个并行化的切片(遍布于节点之间);另一个 RDD 的转换形式;并且最终会彻底改变现有 RDD 的持久性,比如请求缓存在内存中。

4) Storm 流式计算框架

Map-Reduce 计算模型打开了分布式计算的另一扇大门,极大地降低了实现分布式计算的门槛。有了 Map-Reduce 架构的支持,开发者只需要把注意力集中在如何使用 Map-Reduce 的语义来解决具体的业务逻辑,而不用头疼诸如容错、可扩展性、可靠性等一系列问题。但试图用 Map-Reduce 计算模型来解决流处理想要解决的问题时,发现改良 Map-Reduce 并不能使之适应流处理的场景,必须发展出全新的架构来完成这一任务。另一方面,人们对传统的 CEP 解决方案心存疑虑,认为其非分布式的架构可扩展性不够,不具有可扩展性来满足海量的数据处理要求。Storm 的诞生就是针对这个特殊的问题域。

从数据源特征角度来看,Storm 计算框架与 Map-Reduce 明显不同的是:Storm 的数据源是动态的,即收到一条便处理一条,而 Map-Reduce 的数据源是静态的,即数据被处理前整个数据集已经确定,且计算过程中不能被修改。Storm 可以用于 3 种不同场景:事件流(Stream Processing)、持续计算(Continuous Computation)以及分布式 RPC(Distributed RPC)。实时计算框架能够解决很多实际应用问题,如广告推荐、用户行为日志实时分析等。

Storm 体现了以下这些特点:

(1) 可扩展性:当一个集群的处理能力不够用的时候,只要往里面再追加一些新的节点,计算有能力迁移到这些新的节点来满足需要。可扩展性的关键点,即计算所需要的各种状态都是自满足的,不存在对特定节点强依赖,这样,计算就可以很容易地在节点间迁移,整个系统计算能力不够用的时候,加入新的节点就可以了。Storm 的计算模型本身是扩展好的,可以很容易地分布在多个节点上。

(2) 系统可靠性:Storm 这个分布式流计算框架是建立在 Zookeeper 的基础上的,大量系统运行状态的元信息都序列化在 Zookeeper 中。这样,当某一个节点出错时,对应的关键状态信息并不会丢失,换言之 Zookeeper 的高可用保证了 Storm 的高可用。

(3) 计算的可靠性:分布式计算涉及多节点/进程之间的通信和依赖,正确地维护所有参与者的状态和依赖关系,是一个最为关键的问题。Storm 实现了一整套机制,确保消息会被完整处理。

4.4.2 数据挖掘技术

面向智慧城市数据的深度分析、预警与预测,需要借助数据挖掘技术实现。我们需要提供包括聚类、分类、关联规则、协同过滤等机器学习算法的统一注册、应用与注销管理,用于针对特定数据集的挖掘分析。

基于机器学习算法从海量智慧城市数据中提取潜在价值数据,提供面向相关性、推荐、建议、预警等应用分析服务。挖掘分析过程包括数据准备、数据开采、结果表达和解释三个阶段:

(1) 数据准备:具体包含数据集成、选择和预处理三个方面。数据集成将云存储环境中的事实数据进行合并处理,解决语义模糊性、处理数据中的遗漏和清洗脏数据等;数据选择的目的是辨别出需要分析的数据集合,缩小处理范围,提高数据挖掘的质量;数据预处理是为了克服目前数据挖掘工具的局限性。

(2) 数据开采:基于具体的机器学习算法,通过对数据进行分类、度量处理发现新知识。

(3) 结果的表达和解释:根据医疗数据挖掘业务要求,对提取的信息进行分析,把最有价值的信息区分出来,并借助图表等可视化工具展现。为了能更真实反映分析效果,往往需要重复上述过程。

1) 大数据下的频繁模式挖掘技术

频繁模式挖掘是数据挖掘领域一项最重要的技术。频繁模式挖掘搜索给定数据集中反复出现的联系。在传统数据集中,频繁模式挖掘技术已经得到了充分的研究,经典算法有 Apriori 算法、FP-Growth 算法、垂直数据格式数据挖掘算法、闭频繁项集挖掘算法以及这些算法的改进算法和变种算法。

对于智慧城市中的大数据,传统的频繁模式和关联规则挖掘技术将遇到很多问题:首先,由于频繁 $k+1$ 候选集的生成使用频繁 k 项集的自连接生成,因此大数据产生的候选集数量十分巨大,其增长速度是指数级的;其次,为了计算得到所有的频繁项集,算法必须反复迭代搜索数据库,导致高额的 I/O 开销。

考虑到以上问题,并行化的挖掘技术是最佳选择。现有的并行算法包括 CD(Count Distribution)、CAD(Candidate Distribution)和 DD(Data Distribution),这些算法的缺点是高昂的通信开销以及同步问题。Map-Reduce 框架在解决以上问题时具有优势,首先 Map-Reduce 框架能够自动解决失效,其次 Map-Reduce 将程序开发者从复杂的编程中解放出来并提高系统的容错能力。Map-Reduce 提供简化的应用模型,在分布式的环境里提高 Apriori 算法性能,将一个庞大的问题自动切割成小问题,在不同节点并行执行。

2) 大数据下的聚类分析方法

将物理或抽象对象的集合分成相似的对象类的过程称为聚类。聚类是一种重要的人

类活动,也是目前应用最广泛的数据挖掘技术,同时也是一个富有挑战性的研究领域。传统聚类技术主要有基于划分的K均值算法和K中心点算法;基于层次的算法包括利用层次方法的平衡迭代规约的BIRCH算法、分类属性的层次聚类ROCK算法和利用动态建模的Chameleon算法;基于密度的算法包括基于高密度联通区域的DBSCAN算法、通过点排序识别聚类结构的OPTICS算法和基于密度分布函数的DENCLUE算法;基于网格的算法包括统计信息网格STING算法、利用小波变换的Wave-Cluster算法;基于模型的算法有期望最大化方法、概念聚类和神经网络方法;高维聚类算法有维增长子空间聚类CLIQUE算法和维规约子空间聚类PROCLUS算法。

对于智慧城市大数据集,传统聚类算法有很大的不足。根据最近的调查分析,现存的聚类分析算法应用于太或者拍数量级的数据集是不实际的,因为目前最大的单个磁盘容量为1~2 TB,而仅仅读1 TB的数据(5 GB/min 传输速率,eSATA 磁盘)就需要超过3 h。现有的聚类算法是串行的,虽然不适合并行环境,但是作为并行算法在每个磁盘上的“插件”是一种不错的技术选择。

由于目前磁盘容量的限制,将数据存于多个磁盘是必然的。对于智慧城市大数据,目前的技术瓶颈是I/O开销和网络开销。对于算法的设计主要有:算法的设计是否考虑到磁盘传输延迟和网络传输延迟两方面,并能够自动平衡达到最佳运行效果;算法的有效性,由于并行聚类算法不同于串行算法,局部聚类结果不同于全局聚类,因此算法是否可靠有效是一个重要问题;另外,算法可扩展性也是性能的一个重要指标。

3) 大数据下的分类和预测算法

分类和数值预测是预测问题的两种主要类型。目前主要算法有用决策树归纳分类、贝叶斯分类、基于规则的分类、用后向传播分类、支持向量机、近邻学习分类、线性 and 非线性回归等方法。大数据下的预测算法工具有谷歌提供的Google Parallelism R Package。这个框架允许用户使用高级语言编写程序完成将数据拆分并分发到不同谷歌数据中心机器上的任务。

4) 大数据下的序列挖掘算法

序列挖掘主要包含时间序列挖掘、交易序列挖掘和生物序列挖掘。时间序列挖掘是数据挖掘技术的一个重要部分,主要有趋势分析和相似性搜索。趋势分析包括长期运动、周期变化、季节性运动和不规则运动。趋势分析的算法包括自相关分析、 n 阶加权移动平均等。时间序列的相似性搜索(包括时间序列相似性查询)分为全序列匹配和子序列匹配两种方式:

(1) 查询序列和被查询序列的长度相同,称为全序列匹配。

(2) 在较长的时间序列中,找出与查询序列相似的子序列,称为子序列匹配。

根据查询标准,时间序列相似查询可以分为两类:

(1) 最近邻查询,找出 k 个与查询序列最相似的序列或子序列。

(2) 范围查询,找出与查询序列的距离不大于允许误差的序列或子序列。

序列模式挖掘是挖掘频繁出现的有序事件或子序列,经典算法有 GSP、SPADE 和 Prefix-Span 算法。生物学序列比对主要用到马尔科夫模型、隐马尔科夫等。

序列挖掘是一种经常被用于并行化数据的挖掘算法,包括 Monte Carlo 路径模拟等。但是大数据下,目前的算法仍然不能满足需求。下一代的 DNA 序列数据将以空前的速度增长,BLAST 是目前最常用的序列挖掘系统,科技人员已经着手于研究基于 Map-Reduce 的 Blast-Reduce。

4.4.3 大数据可视化技术

数据可视化主要旨在借助于图形化手段,清晰有效地传达与沟通信息。但是,这并不意味着数据可视化就一定因为要实现其功能用途而令人感到枯燥乏味,或者是为了看上去绚丽多彩而显得极端复杂。为了有效地传达思想观念,美学形式与功能需要齐头并进,通过直观地传达关键的方面与特征,实现对于相当稀疏而又复杂的数据集的深入洞察。然而,设计人员往往并不能很好地把握设计与功能之间的平衡,从而创造出华而不实的数据可视化形式,无法达到其主要目的,也就是传达与沟通信息。

数据可视化的成功应归于其背后基本思想的完备性:依据数据及其内在模式和关系,利用计算机生成的图像来获得深入认识和知识。其第二个前提就是利用人类感觉系统的广阔带宽来操纵和解释错综复杂的过程、涉及不同学科领域的数据集以及来源多样的大型抽象数据集合的模拟。

大规模数据的可视化和绘制主要是基于并行算法设计的技术,合理利用有限的计算资源,高效地处理和分析特定的数据集的特性。很多情况下,大规模数据可视化的技术通常会结合多分辨率表示等方法,以获得足够的互动性能。在面向大规模数据的并行可视化工作中,主要涉及 4 种基本技术:

(1) 数据流线化(Data Streaming):将大数据分为相互独立的子块后依次处理。在数据规模远大于计算资源时是主要的一类可视化手段。它能够处理任意大规模的数据,同时也可能提供更有效的缓存使用效率,并减少内存交换,但通常这类方法需要较长的处理时间,不能提供对数据的交互挖掘。离核渲染(Out-of-Core Rendering)是数据流线化的一种重要形式。

(2) 任务并行化(Task Parallelism):把多个独立的任务模块平行处理。这类方法要求将一个算法分解为多个独立的子任务,并需要相应的多重计算资源。其并行程度主要受限于算法的可分解粒度以及计算资源中节点的数目。

(3) 管道并行化(Pipeline Parallelism):同时处理各自面向不同数据子块的多个独立的任务模块。对于任务并行化和管道并行化两类方法,如何达到负载的平衡是关键难点。

(4) 数据并行化(Data Parallelism):将数据分块后进行平行处理,通常称为单程序多数据流(SPMMD)模式。这类方法能达到高度的平行化,并且在计算节点增加的时候可以达

到较好的可扩展性。对于非常大规模的并行可视化,节点之间的通信往往是制约因素。提高数据的本地性也可以大大提高效率。以上这些技术往往在实践中相互结合,从而构建一个更高效的解决方法。

可视化中图形的绘制是一个计算密集型的处理工作。在处理大规模数据时,使用可视化算法,以互动的速度来绘制图形已经超出了单一的 CPU 和 GPU 图形加速器的计算能力。数据并行绘制方法被普遍地用于提高可视化系统的交互速度。最普遍应用的并行绘制算法的分类是基于绘制流水线中图元排序的位置。

近年来受到关注的一种针对模拟计算产生的超大规模数据的可视化模式被称为原位可视化(In Situ Visualization)。它通过将模拟计算和可视化紧密结合,降低数据传输和存储的成本。通常的可视化模式将 PB 量级模拟产生的全部数据传递到存储设备,再经处理后用于可视化。数据传输是整个系统的瓶颈,I/O 将占据绝大部分的计算时间。而在原位可视化模式中,数据直接在计算后原位被缩减与前处理,再用于随后的可视化与分析。经过缩减后的数据,通常比原始数据小多个数量级,能够极大地降低数据传输和存储的开支。

4.5 智慧城市大数据决策分析

从一个城市范围看,智慧城市的核心是智慧。它要求将从不同来源获得的信息,通过快速、专业、智能的分析处理,给出一个可度量的、可评估的科学决策方案。智慧城市大数据决策分析架构如图 4-8 所示。

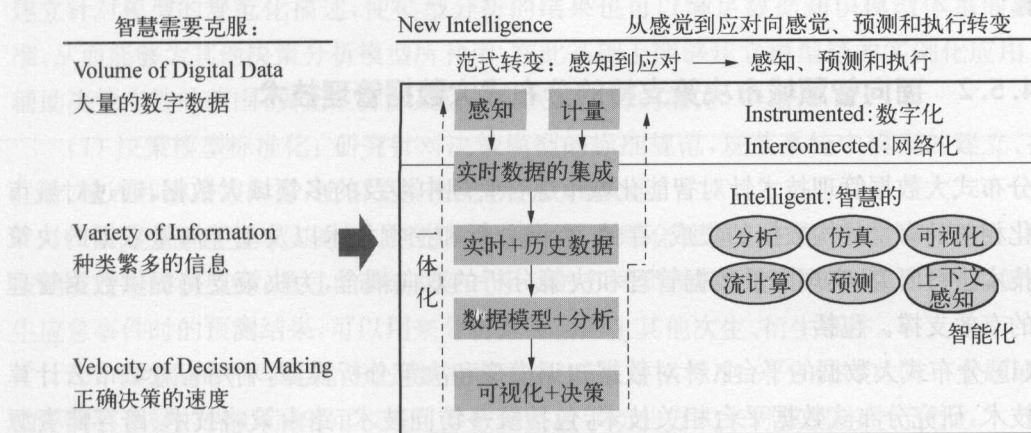


图 4-8 智慧城市大数据决策分析架构

智慧城市中大数据相关的关键技术包括融合多领域信息的数据知识模型、面向智慧城市决策支持的分布式海量数据的管理技术、辅助决策模型的智能组合技术、高性能多用户

协同的辅助决策技术等。

4.5.1 融合多领域信息大数据的知识模型

融合多领域信息大数据的知识模型,向与城市综合管理相关的各领域(如交通、卫生、食品、环境、能源、规划等)数据提供统一的数据采集、整理、处理等支撑服务。其次,对城市综合管理决策可能涉及的多领域的需求,在标准统一的领域数据知识模型的基础上,建立面向不同主题的多领域数据知识模型,如,疾病的发生可能与环境变化、食品污染、药品短缺等相关,食品价格的波动可能与交通物流状况、环境带来的产量变化等相关。最后,在建立面向领域和面向主题(涉及多领域)的数据知识模型的基础上,形成多领域多层次数据知识模型结构。相关技术包括:

(1) 数据和知识表达形式的规范性:在对交通、卫生、食品、环境、能源、规划等领域知识研究的基础上,提出科学、统一、规范的知识信息表达形式,为进行多领域综合分析提供统一的数据基础。

(2) 面向领域的知识模型:通过对交通、卫生、食品、环境、能源、规划等不同领域的关键业务和数据类型进行研究,建立面向不同领域的、规范的数据知识模型。

(3) 面向主题(多领域)的数据知识模型:在规范的数据和知识表达形式的基础上,研究面向多领域的综合分析业务需求,建立面向主题(多领域)的综合数据知识模型,如流行病分析、价格波动分析、社会稳定分析、城市规划决策等主题。

(4) 多领域多层次数据知识模型结构:在面向领域和面向主题的数据知识模型的基础上,研究多层次的、统一的数据知识模型,为建立智能化城市综合管理辅助决策模型提供技术支撑。

4.5.2 面向智慧城市决策支持的分布式大数据管理技术

分布式大数据管理技术针对智能化城市综合管理中涉及的多领域大数据,通过对城市信息化相关的海量异构数据的集成、存储、查询分析和挖掘技术以及基于海量数据的决策分析技术进行研究,实现海量数据管理和决策分析的双向耦合,为决策支持提供数据管理层面的有效支撑。包括:

(1) 分布式大数据的平台:针对数据知识模型和决策分析模型,利用智慧城市云计算相关技术,研究分布式数据平台相关技术,包括统一访问技术、路由策略技术、跨存储资源访问技术、事务技术等,建立面向智能化城市综合管理的分布式数据平台,提供高效的数据组织、存储和访问的机制。城市管理中涉及大量的实时数据应用,这些应用使得数据量呈现快速增长的趋势。另一特点是存在突发性的数据爆发。因此,需能够支持存储平台的负载均衡和弹性伸缩,主要包括云平台通用的负载均衡和弹性伸缩技术、模型数据的分布式

存储和基于查询频率的负载均衡技术。

(2) 基于云计算的海量异构数据的存储管理技术: 城市管理原始数据包含大量的结构化、半结构化和非结构化数据。模型相关数据包含多种表示方法: 向量数据、序列数据、图数据。目前的云数据服务技术主要针对结构化数据和基于 Key-Value 的文本数据, 而对于序列数据、图数据这些类型没有很好的支持技术。项目将基于 Hbase、HIVE 等 NoSQL 云数据库系统, 研究 Hadoop 平台上数据管理与访问的技术, 主要包括面向多维度聚集查询的数据的分割与分布策略、大规模聚集查询的分布式实施方法、以流数据为代表的海量监测数据的快速写入与基于 (Key, Value) 方式的分布式存储与访问方法、面向曲线特征的监测数据组织方法。

(3) 分布式决策模型: 研究面向海量异构数据的智能城市的决策模型。在分布式计算框架 Map-Reduce 的基础上, 对决策模型 (多元回归、神经网络、主成分分析等) 进行研究, 使其具有分布式计算的能力, 能够对分布环境下的数据库进行分析, 以提高算法的运算能力。

(4) 流数据挖掘技术: 智能城市分析的另一特点是数据的快速性, 因此需利用流数据分析和挖掘算法, 构建快速实时的决策分析模型。针对智能化城市综合管理中的海量实时数据“连续、有序、变化、快速、海量”的特点, 采用分类、聚类、频繁模式挖掘等方法来处理城市不同领域的监测数据。

4.5.3 辅助决策模型的智能组合技术

智能化城市综合管理中, 有多种面向领域和面向主题 (多领域) 的决策分析模型, 这些决策模型的输入, 不仅有获取的城市管理基础数据, 而且有其他模型的分析结果, 因此需要建立针对模型的规范化描述, 使模型分析的结果也可以满足数据知识模型体系的数据标准, 从而能够为其他决策分析模型所利用, 在此基础上能够建立模型链的实例化应用, 扩展辅助决策支持的范围, 真正发挥多层次辅助决策支持的能力。包括:

(1) 决策模型标准化: 研究针对决策模型的标准规范, 规范系统中模型的建立、描述、注册、输入和输出等内容, 使相关模型能够组成模型链。

(2) 模型链建立与实例化: 通过对智能化城市综合管理业务需求的分析, 建立对应某主题或过程的模型链, 通过自动或交互式手动的方式实现模型链的实例化和执行, 例如, 发生应急事件时的预测结果, 可以用来分析是否会产生其他次生、衍生事件。

(3) 模型链评估与优化: 通过建立模型链比较、评估标准, 来分析针对同一主题或过程的不同模型链的优劣, 同时可以通过参数调整或手动的方式来对模型链进行优化。

4.5.4 高性能多用户协同的辅助决策技术

智能化城市综合管理中积累了各领域的海量历史数据, 对这些数据进行综合分析能够

提供智能辅助决策的功能。很多智能辅助决策算法的计算结果与地理信息、空间信息关系紧密,通过可视化技术,以更直观、有效的方式将分析结果展示给用户,同时,将分析结果按时间切片顺序进行展现,能够模拟出动态演变发展的过程,显示出不同现象与动态变化之间的关系。主要包括:

(1) 决策模型的可视化展现技术:将决策模型与具体的业务需求相结合,对实时动态信息或历史信息进行分析,分析结果与GIS、虚拟现实环境相结合,能够进行基于场景的可视化展现,如危险气体扩散模型、地铁客流量预测模型等。

(2) 可视化动态分析模型:以时间切片为依据,将决策模型的可视化结果顺序地展现给用户,使用户对整个的发展演变过程有最直观的认识,既可以是历史数据的分析,也可以是对事态发展的预测。同时,用户可以通过调整决策模型的参数、最优分析评估等手段,来获得不同的事态发展过程展示,提高系统的智能化水平。如:危险气体在不同条件下的影响范围趋势,地铁客流量在不同站点、不同时间段条件下的预测分析等。

(3) 可视化智能决策方案:在对获取的历史和动态信息进行分析的基础上,系统结合业务知识,自动生成当前信息的可视化决策方案,同时,用户可以通过参数调整等手段更新决策方案。如危险气体泄漏事件中的可视化智能救援方案、基于历史信息的城市优化规划决策方案等。

◇ 参 考 文 献 ◇

- [1] Dan P. BASE: An Acid Alternative. <http://queue.acm.org/detail.cfm?id=1394128>.
- [2] Julian B. Brewer's CAP Theorem. <http://www.julianbrowne.com/article/viewer/brewers-cap-theorem>.
- [3] Sanjay G, Howard G, Shun-Tak L. The Google File System. <http://research.google.com/archive/gfs-sosp2003.pdf>.
- [4] Fay C, Jeffrey D, Sanjay G, et al. Bigtable: A Distributed Storage System for Structured Data. <http://research.google.com/archive/bigtable-osdi06.pdf>.
- [5] Jeffrey D, Sanjay G. MapReduce: Simplified Data Processing on Large Clusters. <http://research.google.com/archive/mapreduce-osdi04.pdf>.
- [6] Anand. Rajaraman, Jeffrey. David. Ullman. 大数据: 互联网大规模数据挖掘与分布式处理[M]. 王斌,译. 北京: 人民邮电出版社,2012.

第5章

智慧城市中的大数据综合应用

智慧城市助推大数据的产生,也给大数据提供了广阔的应用市场。智慧城市的基本特征是“三化一新”,即物联化、互联化、智能化和创新。物联化即充分应用物联网手段,感知城市各主体运行的实际状态,包括城市土地、水、空气等自然资源,道路、楼宇等各类人造公共基础设施,交通控制等各行业系统以及以人为中心的社会系统。互联化是实现智慧城市各系统的互联互通,促进数据共享,减少信息孤岛。智能化是通过对各行业数据的充分应用,提升各行业运作效率。创新是在数据共享和智能应用的基础上,在城市的社会管理和公共服务领域引入新模式,提升城市综合竞争力,促进城市转型发展。智慧城市大数据采集与应用如图 5-1 所示。



图 5-1 智慧城市大数据采集与应用

智慧城市中的核心是人,人在城市生活中会产生大量的数据。医疗卫生是与人们生活最密切的行业,医疗数据是智慧城市大数据的重要构成部分。上海医联工程整合了上海 3 000 多万人的健康档案和诊疗记录,2013 年仅处方数据就超过 20 亿条。

智慧城市需要产业支撑,尤其是依托新科技的创新产业。新产业的发展产生大数据并

受益于大数据。腾讯、百度、阿里巴巴这样的互联网巨头,带动了互联网产业的巨大发展,也积累了丰富的数据资源。以“双 11”购物狂欢为例,据国家邮政局 2013 年 11 月 12 日公布的信息显示,11 月 11 日全天,网络消费共产生快递订单约 1.8 亿件,预计各大快递企业每家至少需处理 6 000 多万件包裹。1 天就可以产生数亿条数据记录,这充分说明了目前的智慧产业有着明显的大数据特征。

应用大数据能提升城市智慧。通过对数据的分析,可以形成趋势判断、可以建立不同事件之间的关联,从而提升城市智慧。对医疗大数据的充分应用不仅能帮助政府更科学地进行医保预算,更智能地实现医保控费,还能建立多种疾病之间的关联,提升医疗机构的疾病诊断能力,提升市民的健康水平。应用电子商务大数据对于预测经济状况、按需生产、提升城市智慧非常有帮助。例如,跟进对大量的消费数据进行分析、合理安排物流配送点,对降低物流成本和提升消费满意度非常重要。

智慧城市需要打破数据孤岛,解决大数据应用中的共性问题。传统以部门为中心的城市信息化形成了不少数据孤岛。智慧城市中的大数据采集、共享与分析利用牵涉多个行业,如果每个行业都独立去做,大数据孤岛的危害会更大,城市需要一体化的思路,处理各行业在大数据应用中的共性问题。

以城市安全和社会管理为例,公安、国安等安全部门掌握着大量的数据,包括视频监控图像、情报信息等,其他各级政府部门也掌握着大量的数据,如交通、城管、街道、居委等,此外还有各类社会资源,例如大型商场、楼宇、公共场所的数据。这些部门往往各自管理和使用这些数据,不能发挥数据协同的效应。例如一个突发事件发生,如果要追踪嫌疑人员,往往只能通过人工方式对每个单位的视频进行逐个解读,耗时费力。智慧城市可以建立起统一的视频大数据管理系统,布局对城市各机构视频大数据整合,为相关单位提供智能的视频分析。

本章主要包括:网络数据采集平台,跨部门大数据管理平台,大数据开放服务平台和大数据决策支持平台。其中,网络数据采集平台主要介绍互联网数据采集、物联网数据采集和个人数据采集;跨部门大数据管理平台主要介绍城市运营大数据、城市一体化公共信息平台和大数据平台服务。

5.1 网络大数据采集平台

大数据的价值不仅仅在于其数量庞大,关键是其网络化特征核心。只有能够被网络化利用的数据,才可能成为大数据,并在城市特定领域发挥价值。信息社会进入大数据时代,网络大数据是主体,网络大数据采集也是智慧城市重要的数据来源。网络大数据除了大数据的一般特征外,多源异构以及垃圾数据多是非常重要的特征,建立网络数据采集平台对于简化智慧城市的大数据采集非常重要,这能让智慧城市的各个系统聚焦在大数据的应用

开发上,而不是在多源异构网络大数据采集上重复投入。

5.1.1 互联网数据采集

互联网让信息更容易获得,搜索引擎爬虫更是将互联网变成一个巨大的数据库。20 年的互联网发展积累了海量数据,并且还在以更快的速度增长。据 IDC 报告,预计 2020 年,全球数据总量将 40 ZB(4 万亿 GB),是 2011 年(1.8 ZB)的 22 倍^[1]。全球数据总量中,1/3 是互联网数据。例如 eBAY 的数据已经超过 100 EB。以下是一些互联网统计数据。

- 6.34 亿——全球网站总数(截至 2012 年 12 月)。
- 150 亿——Tumblr 月浏览量,每天 3 TB 的数据存储。
- 560 亿——美国社交新闻网站 Reddit.com 年浏览量。
- 1 440 亿——全球每天发送的电子邮件总量。

中国已经进入网络社会,智慧城市的建设将受益于互联网的发展。2014 年 1 月 16 日,中国互联网络信息中心(CNNIC)发布的第 33 次《中国互联网络发展状况统计报告》显示,截至 2013 年 12 月,中国网民规模达 6.18 亿,互联网普及率为 45.8%^[2]。随着互联网的普及,互联网内容和服务还将不断丰富,全球数据量会继续增加,互联网数据会成为重要的资产。

互联网数据的采集涉及页面和后台数据两个层面。就网页而言,HTML 为主的超文本蕴含了丰富的内容,可以通过爬虫程序直接抓取所需的数据。网站后台数据相对层次较深,在数据采集时比页面数据采集要复杂一些。除了网站授权通过相应接口访问外,越来越多的网站开始实施开放策略,通过开放的应用程序接口(Open API)让各方可以很方便地获取数据,开发应用。

考虑到互联网海量的网页,为了保证大数据抓取的实时性,可以建立大数据抓取集群,该抓取集群作为一种公共服务,面向各类抓取需求,提供抓取规则管理,实现统一的互联网数据抓取,如图 5-2 所示。

社交网络已经成为互联网上人气最旺、数据最鲜活的部分。微博、微信等社交网络方便了人们的交往,也留下了人们的行为痕迹,这些行为痕迹的采集和分析,对智慧城市的社会管理和公共服务有着巨大的价值。政府要采集人们的基本信息,尤其是工作、居住变动的信息往往非常困难,成本很高。而人们往往愿意在社交网络,如朋友圈公布一些个人信息,这些信息可能包括搬家、生育、旅行、就业、消费等。智慧城市要在公共服务均等化的基础上,为人们提供个性化、智能化的服务。这是智慧城市多个行业的需求,可以建立起城市一体化的社交网络采集系统,采集的数据包括:居民的个人信息变更、居民消费水平、消费倾向、居民对公共服务的评价、居民社会关系等。印第安纳大学学者利用谷歌提供的心情分析工具,对 270 万用户在 2008 年 3~12 月所张贴的 970 万条留言,挖掘出用户快乐、友善、警觉、确信、活力和冷静等 6 种心情,进而对道琼斯工业指数的变化进行预测,准确率达到 87%。2004 年,央视新闻联播采用百度大数据服务,在两会期间跟进互联网尤其是社交

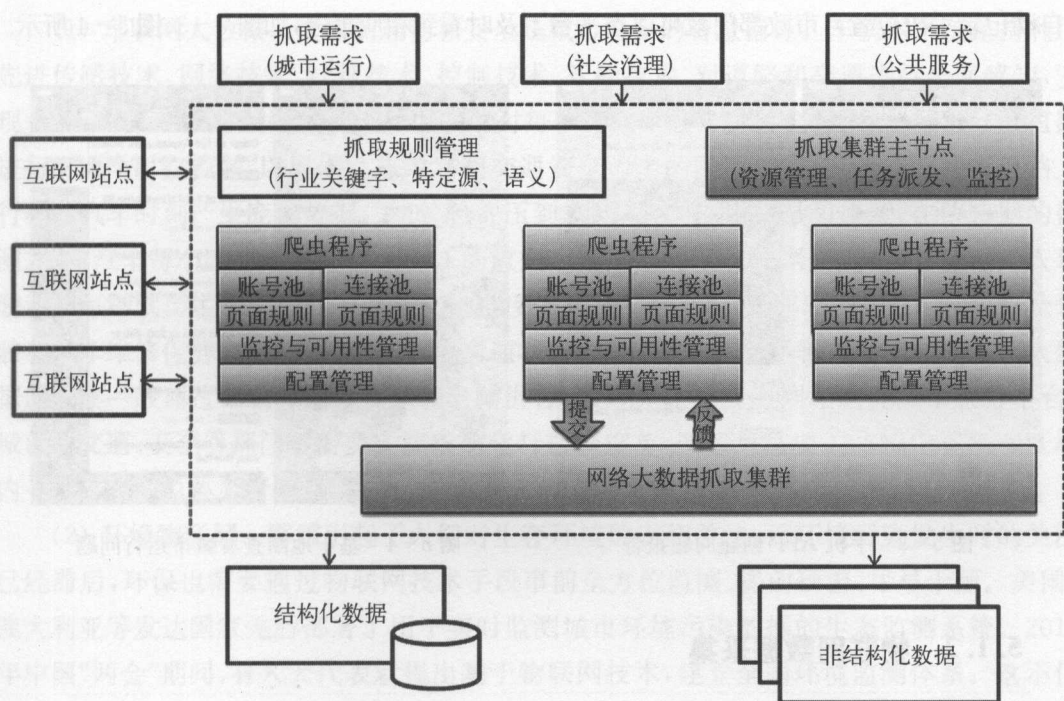


图 5-2 智慧城市互联网大数据抓取

网络大数据,分析了人们对政府的关注点,这对于政府推进新型城镇化、工业化、信息化和农业现代化的建设,尤其是建设智慧城市,实时信息惠民都非常重要。

移动互联网是智慧城市的重要基础设施,也将助推大数据的生产。目前移动互联网的速度已经超过了传统桌面互联网的发展速度,根据工信部发布的数据,中国移动互联网用户在 2014 年 1 月已经达到 8.38 亿。移动互联不仅改变着人们生活方式,也积累了大量的数据。如今,一部普通的智能手机内存就有 1 GB,扩展存储一般也达到了 8~32 GB。根据 IDC 的数据,2013 年全球智能手机的出货量就超过了 10 亿部。这些手机在上网、社交、购物、阅读、记录等方面产生的数据是惊人的。

由于手机等智能终端一般内置全球定位系统(GPS),移动互联网的大数据往往与地理位置是关联的,这对于智慧城市的运行管理和服务价值更大。移动互联网的大数据包括:移动搜索、移动视频、移动社交(微信、微博)、移动支付、移动位置(路径)等。通过手机 APP、轻应用等形式,能很方便地采集移动互联的数据,并对海量用户的移动互联网数据进行统一的分析,为城市运行管理提供趋势判断。在移动互联网数据发布和采集时,采用开放平台和接口,能屏蔽移动互联网设备的技术差异,这对于智慧城市的开放性非常重要。

建立城市统一的移动互联网数据采集平台能够更好地连接政府和市民,让市民通过手机应用参与城市治理。美国波士顿市建立了一个 Citizens Connect 系统,让市民成为城市的眼睛和耳朵,通过手机报告周边的问题,例如漏水、窨井盖丢失、路灯故障等城市基础设施问题,从而提醒市政管理部门及时处理。对于市民而言只需要下载相应 APP、拍照上传

(自动记录手机位置),市政部门就可以在平台上及时看到相应报告,如图5-3、图5-4所示。

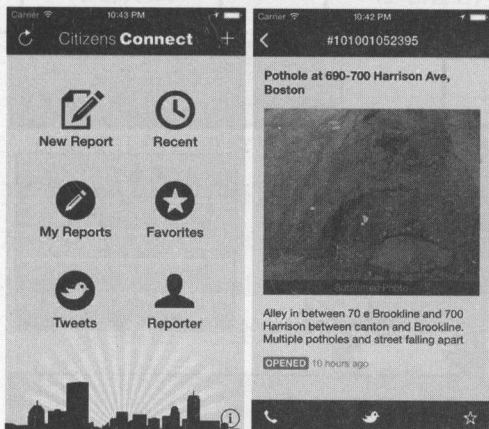


图5-3 手机APP创建问题报告

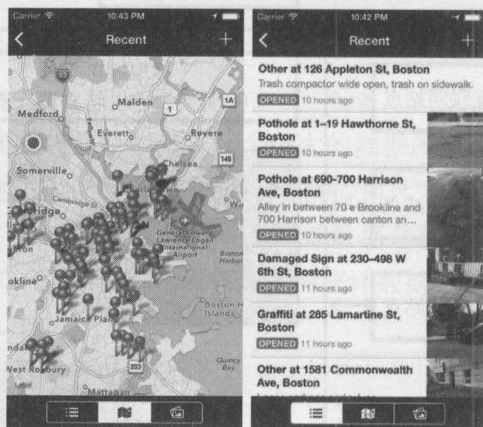


图5-4 基于地图查看城市运行问题

5.1.2 物联网数据采集

物联网在城市各行业的充分应用,让智慧城市有了共同感知的基础。按照维基百科的定义,物联网是一个基于互联网、传统电信网等信息承载体,让所有能够被独立寻址的普通物理对象实现互联互通的网络。物联网让传统的PC-主机网络得以扩展,物-物、人-物、机-物互联成为可能,也为智慧城市提供了更大量的数据。

平安城市视频监控(“天网”)就是很典型的物联网大数据采集。安全是城市最基本的需要。天网工程,也就是城市电子监控系统,是为满足城市治安防控和城市管理需要,利用视频监控技术与传输网络,对城市固定区域进行实时监控和信息记录的视频监控系统。随着物联网技术的反战,平安城市也从传统的视频监控向基于大数据的社会管理转化。对平安城市而言,一体化的城市管理离不开大数据的采集和分析。伦敦的摄像头一天产生的数据量就达8TB。目前国内城市的视频监控一般基于高清摄像头,高清视频的流量可以达到8MB,一个中大城市的摄像头数量超过20万,这些摄像头24小时不间断传送视频流,数据长年累月积累。

对智慧城市而言,没有必要每个行业重复部署、采集、存储和管理这些视频数据,整个城市建议一套标准化的视频采集系统即可,对视频大数据进行统一采集、统一管理,并在视频大数据的基础上,建立起事件索引、事件关联和智能比对等共性功能,各行业按需调用相应的视频大数据服务即可。例如对于节日期间的视频大数据进行智能分析,公安部门可以通过图像智能匹配,寻找特定人员,并标示特定人员在城市中一段时间的行动轨迹,便于关联相关事件;交通部门可以对交通出入口的视频进行人流密集度分析,跟进动态的人流密度,合理调节运输资源。

(1) 车联网大数据。交通的通畅和安全是城市重要的智慧指标。车联网系统是指利用先进传感技术、网络技术、计算技术、控制技术、智能技术,对道路和交通进行全面感知,实现多个系统间大范围、大容量数据的交互,对每一辆汽车进行交通全程控制,对每一条道路进行交通全时空控制,以提供交通效率和交通安全为主的网络与应用。如今,城市道路上行驶的汽车时刻产生庞大数据。从轮胎气压到发动机转速,到油温和速度,在传感器的监测之下,汽车每小时能产生 5~250 GB 的数据。高级车产生的数据更多,如谷歌的无人驾驶汽车每秒就产生 1 GB 的数据。每小时的数据量就超过 3 TB。车联网数据不仅对安全驾驶和汽车维修保养有用,对于分析城市交通状态、交通事故预警都非常有用。车联网大数据的采集一般通过车载传感器和车载电脑进行,通过互联网传送到相应的汽车服务平台,城市的交通、安全等部门不需要关注车辆运行数据细节,一般只需要关注特定车辆和区域内一批车辆的总体运行数据,获得及时预警。例如对于公交车、出租车的监管。

(2) 环境物联网。雾霾引起了人们对生存环境的密切关注,而环境污染发生时的关注已经滞后,环保也需要通过物联网技术手段事前全方位监测、提前预警、尽早干预。美国、澳大利亚等发达国家先后部署了用于实时监测城市环境污染数据的生态监测系统。2014 年中国“两会”期间,有人大代表就提出基于物联网技术,建立全国环境监测体系。这不仅是哪一个城市的任务,也是多城市、多区域协同的任务。基于一定的标准体系,环境传感器采集的大数据能被多部门和社会所共享。如社会力量可以开发天气 APP、空气质量 APP,这些服务要保证数据的准确性,无须自行部署传感器,只需要接入环境大数据平台就可。

5.1.3 个人数据采集

城市中的每一个人都与大数据密切相关,不仅会受益于大数据带来的管理和服务的智能化,也会产生大数据。如今社会,个人产生大数据主要有如下几个方式:

(1) 个人在互联网上生成内容,包括社交网络内容。

(2) 日常生产生活中与企业、资源设施互动中产生数据。

(3) 通过可穿戴设备产生传感数据。

个人数据的采集方式如图 5-5 所示。

相对于专业的互联网内容制作而言,近几年用户生产内容开始兴起。根据维基百科的定义,用户生成内容(User-Generated Content,UGC)指网站或其他开放性媒介的内容由其用户贡献生成。从 2005 年左右开始,互联网上的许多站点开始广泛使用用户

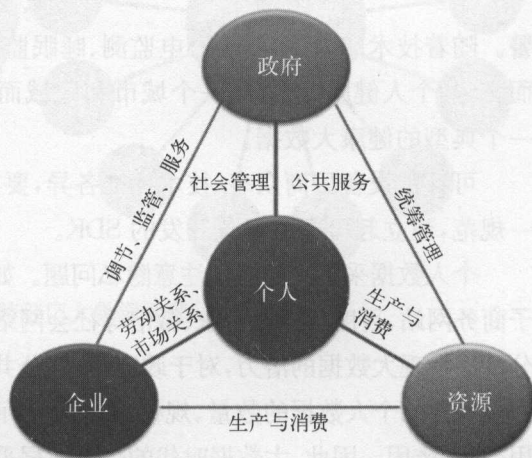


图 5-5 个人数据的采集方式

生成内容的方式提供服务,许多图片、视频、博客、播客、论坛、评论、社交、维基、问答、新闻、研究类的网站都使用了这种方式。

用户生成内容是 Web 2.0 概念的组成部分之一。部分用户生成内容站点也会使用或提供网站的开源、自由软件程序或相关 API 支持,以促进用户的协作、技术支持和对网站的发展。

用户生成内容的意义在于三个方面:一是极大地丰富了互联网数据,而且让网络数据有了生命力,随着新用户的加入而不断完善。二是采集和应用群体用户生成的内容对于研究群体的思维与行为方式,分析和预测特定事件、特定时空、特定群里的行为很有帮助。三是根据特定用户全面的网络生成数据,可以更全面地了解个体,提供更个性化的服务。

个人与城市中的企业、政府和资源设施的互动构成了城市生活和社会系统。人们在日常的生产生活过程中也会产生大量的数据。个人从出生开始,在学习、就业、医疗、饮食、住房、出行、文化娱乐、人际交往、退休、养老等社会生活各方面会生产大量的数据。采集、分析和应用这些数据,对于创新社会管理、提升公共服务非常重要。

可穿戴设备是一种可以穿在身上或贴近身体并能发送和传递信息的计算设备。可穿戴设备自身轻巧、紧贴身体的特点,让其成为身体与世界交流的一个更适合载体。从这个互联网上的定义来看,可穿戴设备作为一个载体,能够实时采集和传递数据为个人以及社会服务。

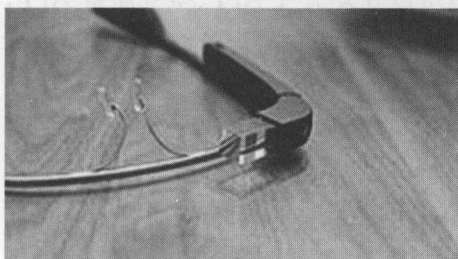


图 5-6 2013 年发布的谷歌眼镜

于 2013 年发布的谷歌眼镜是希望用眼镜取代智能手机的屏幕,并且允许使用自然语言来与互联网交互,谷歌眼镜不仅可以显示,加上摄像头和传感器就是一个随时随地的视频、照片和环境数据采集仪,如图 5-6 所示。

如今,与健康相关的个人可穿戴设备也开始流行。这些设备可以实时采集个人的脉搏、血压、血氧,甚至血糖等体征数据,判断身体状况,及时预警。随着技术的发展,随身心电监测、睡眠监测等专业的移动健康设备也开始应用,就个体而言,是个人健康数据,就一个城市和区域而言,把大量人的健康监测数据汇集起来,就是一个典型的健康大数据。

可穿戴设备厂商众多,技术标准各异,要有效采集和应用这些数据,需要在接口层面统一规范,建立起可穿戴设备开发的 SDK。

个人数据采集需要特别注意隐私问题。如今不仅政府关注和采集人们的数据。淘宝等电子商务网站、百度等搜索引擎、微信等社会网络都在记录和分析人们的数据。通过这些大数据分析能发现大数据的潜力,对于政府管理、公共服务和商业服务都非常有用,这进一步导致了采集和存储个人数据的数量、规模急速增长,而且这些数据甚至是个人隐私数据会被重复利用,甚至滥用。因此,大数据时代的个人数据采集会让人们的隐私受到威胁。

维克托·迈尔-舍恩伯格在《大数据时代》一书中对大数据的风险与掌控进行了深入分

析。他认为,在大数据时代,不管告知与许可、模糊化还是匿名化,这三大隐私保护策略都失效了。他提出了掌控大数据的几点管理变革:个人隐私保护,从个人许可到让数据使用者承担责任;防止滥用大数据预测个人动因,个人应该为他们的行为负责而非倾向负责;击碎黑盒子,让大数据程序员发挥专业价值;反数据垄断大亨^[3]。

5.2 跨部门大数据管理平台

传统城市信息化形成了大量的信息孤岛,多个政府部门都积累了各自领域的的数据,但部门之间数据共享程度低、数据重复采集现象严重。以视频监控数据为例,公安、交通等多个部门都在路口、重点场所安装部署摄像头,分头采集和存储视频监控数据;大型楼宇、公共场馆、商场、小区等机构也都有各自的视频监控系统,这些视频大数据没有共享,突发事件发生时,往往要靠人工走访各部门,获取视频数据,依赖人工建立多个视频数据的关联。

智慧城市要实现智能化管理和一体化服务,就需要解决城市信息孤岛问题,建立起跨部门的数据共享体系。跨部门的大数据管理平台以城市一体化的视野,为各领域的智慧应用提供大数据统一服务,如图 5-7 所示。

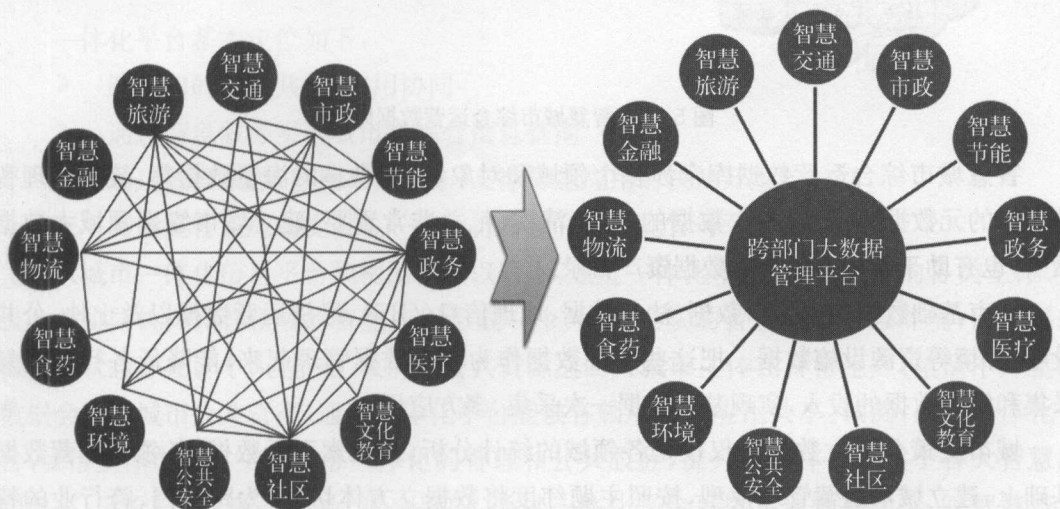


图 5-7 智慧城市跨部门大数据管理

5.2.1 城市运营大数据

城市运营大数据包括安全、交通、医疗、环境等多行业专题大数据,为了更好地实现城

市级的大数据管理和服 务,建立智慧城市综合运营数据库(City Operational Database, CODB)实现城市一体化的大数据管理很有必要。

CODB 是智慧城市的大数据集散中心,是各智慧行业应用数据交换、共享的枢纽,按请求的领域、行业的层次、粒度来进行数据路由。

CODB 将由各智慧行业应用库的索引元数据库、城市基础数据库和行业综合库三大信息库构成,如图 5-8 所示。

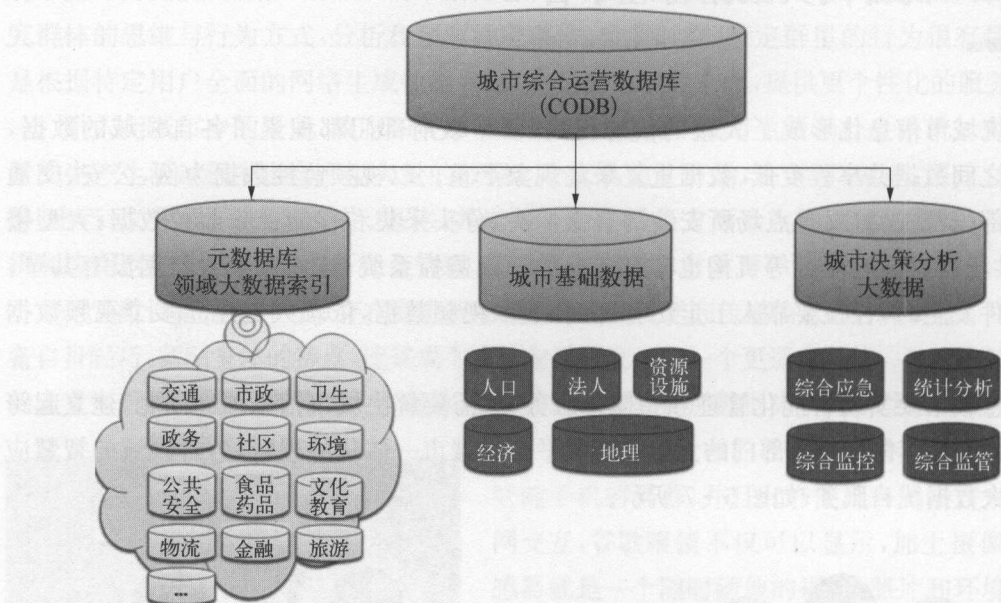


图 5-8 智慧城市综合运营数据库

智慧城市综合运营数据库牵涉多个领域和对象,拥有众多的海量数据集,统一管理智慧城市的元数据,对于城市大数据的抽取、清洗、汇总非常重要,建立城市级别领域大数据索引,也有助于管理好城市大数据资产目录。

城市基础数据包含人口数据、法人数据、地理信息(GIS)、经济运营数据以及土地、公共设施、环境等资源设施数据。把这些共性数据作为基础数据管理起来,能降低各行业重复采集和维护数据的投入,实现基础数据一次采集、多方应用。

城市决策分析大数据不仅仅是各领域的统计分析,还在索引元数据、各领域运营数据基础上,建立城市运营管理模型,按照主题纬度将数据立方体切片,为跨部门、跨行业的特殊目标主题应用服务。例如对城市运营状态的综合监测,分析城市安全指数、健康指数、交通指数、生活指数等,为处置突发状态提供智能预警和及时反馈。

智慧城市大数据按照全生命周期原则进行统筹管理,按数据性质分级管理,按数据类型规划存储方式,按数据粒度规划存储介质。

5.2.2 城市一体化公共信息平台

数据和应用是智慧城市系统的核心。为了各智慧应用之间的共享与协同,智慧城市需要构建一体化公共信息平台(简称一体化平台),实现跨部门的信息资源管理,并利用大数据思维把分散在各领域的的数据管理起来,并对外服务,如图 5-9 所示。

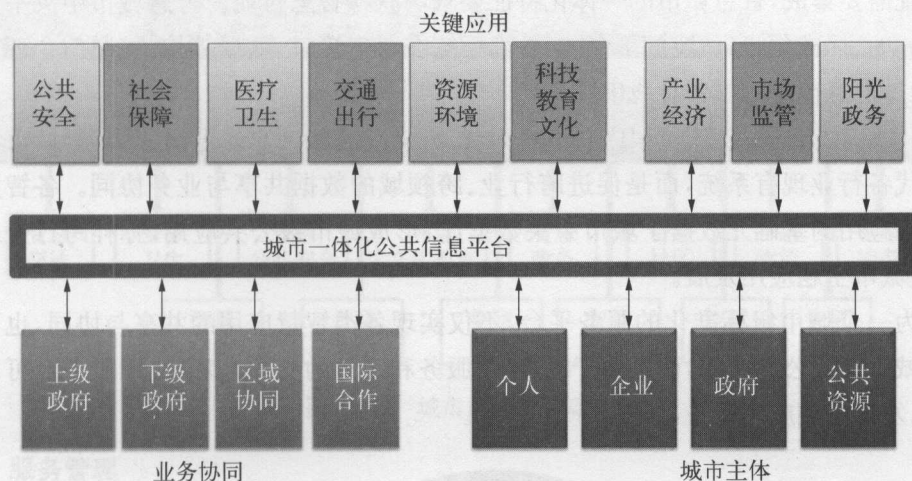


图 5-9 城市一体化信息平台

一体化平台基本定位如下：

- 跨部门的信息共享、应用协同
- 通过信息服务实现城市级综合信息查询
- 通过数据服务、应用支撑服务,支撑智慧应用的快速开发与运行
- 为各类智慧应用提供统一的运行管理和应用托管等平台服务

从城市一体化信息平台架构图中可以看出,通过一体化平台,智慧城市的各类主体、各级业务协同机构和各领域的智慧应用形成一个开放、互联的智慧城市生态环境,而不再以孤岛的状态存在。个人、企业、政府和公共资源这四大城市主体在日常生产、生活中产生的数据会成为城市公共资产,通过一体化平台能被各领域智慧应用共享、调阅;通过一体化平台,城市主体也能够实现城市一体化的管理和公共服务,提升服务体验。对于各大智慧应用而言,无论归属于哪个行业、哪个部门,通过一体化平台可以实现数据共享,对智慧城市的各类数据资源进行统一的查询、订阅和发布。

1) 平台功能

(1) 城市各系统数据共享: 各智慧应用的数据交换与共享;基于事件触发的智慧应用协同;基础设施服务的平台化,按需分配和回收机制,计算、存储、网络资源的复用,安全服务的协同。

(2) 数据统一管理分析: 依托城市中央数据库和各行业数据库的集成, 实现城市信息资源的分布与集中并存, 元数据和基础数据的统一, 主题数据切片, 数据的循环生命周期管理(生产、采集、过滤、汇总、挖掘、发布、再利用、销毁)等。

(3) 快速应用服务支撑: 快速应用服务的提供依赖成熟的支撑平台和模块化组件。通过提供信息服务、应用服务等基础的服务组件, 能通过服务组合、应用聚合快速发布新的应用服务。

(4) 第三方应用集成能力: 随着物联网技术在智慧城市中的深入运用, 专业外设、外部专用系统需要集成, 智慧城市的一体化特征要求串联各智慧行业。智慧城市中央平台应是一个开放的、标准化、自适应的平台, 与外部系统在通信接口、数据结构、控制信令环节无缝融合, 通过消息传递的服务总线串联。

(5) 城市系统平滑演进能力: 城市云中心是城市信息化的演进而非颠覆, 城市云中心并非取代各行业现有系统, 而是促进跨行业、跨领域的数据共享与业务协同。各智慧行业汇聚业务应用的基础元数据于城市中央数据库, 形成城市级公共应用, 弥补跨行业应用空白, 促进城市主题应用发展。

作为一个城市级标准化的服务平台, 不仅实现各类智慧应用的共享与协同, 也可以直接承载城市级的公共应用, 而且由于综合库、服务和工具的支撑, 多家应用服务商可以方便地开发、发布各类应用服务, 如图 5-10 所示。

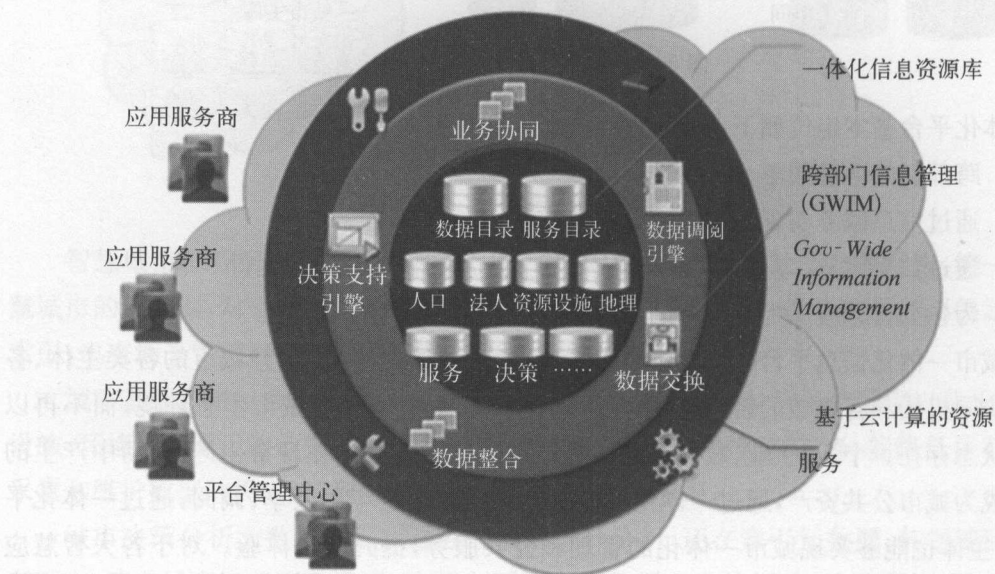


图 5-10 城市级标准化的服务平台

2) 城市信息共享架构

智慧城市是一个复杂的巨系统, 城市一体化平台是系统之系统(System of Systems), 需要有很强的兼容性。在设计一体化公共信息服务平台时采用多种架构的融合, 包括面向服务的架构(SOA)和面向资源的架构(ROA)等, 如图 5-11 所示。

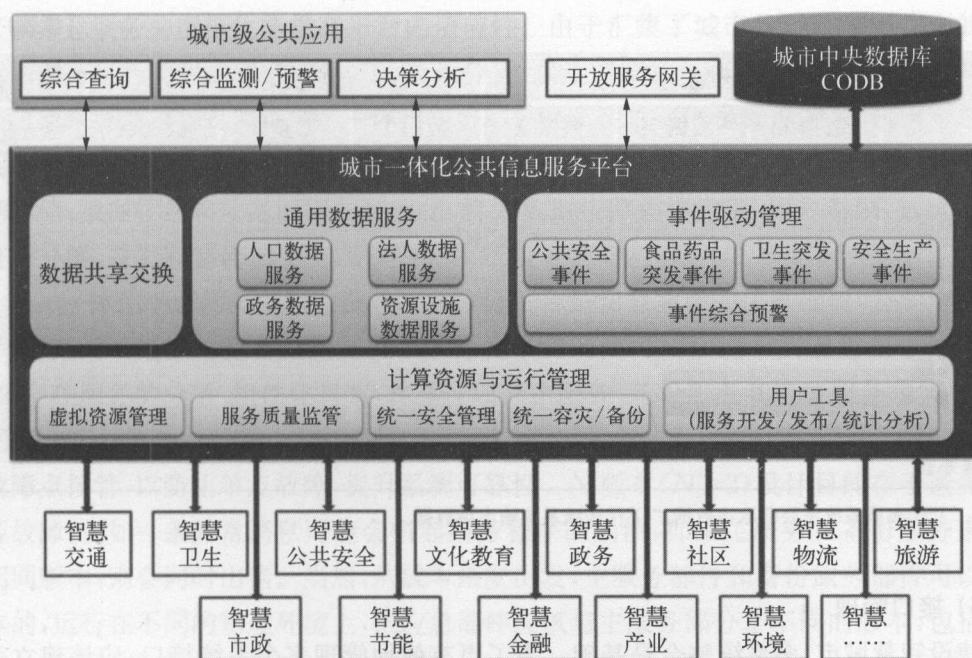


图 5-11 城市信息共享架构

3) 服务管理

把信息共享和应用协同需要的各类基本服务组件化封装,作为一体化平台可配置、可运行、可管理的服务组件。这些服务组件包括通用技术组件和城市关键业务组件。通用技术组件如数据交换服务和应用支撑服务组件;关键业务组件如人口、法人等基本信息调阅,让各类跨部门应用可以方便、安全、规范地访问中央数据库的内容。

4) 事件管理

城市的日常运行可以分解为各类事件,这些事件包括常态化的事件,也包括紧急事件。一体化平台通过订阅发布、广播等机制形成跨部门的城市级大联动效果。例如,一个食品安全事件发生,平台可以根据预先存储的规则,判断出需要联动的系统,并自动将信息发送给这些系统,这些系统的用户可以在第一时间得到这些信息,而无须人工干预。

城市运行事件管理不仅是事件的订阅和推送,还包括相应数据智能推送,对于智慧城市而言,可以建立起如表 5-1 所示的系统/事件矩阵,更好地实现跨部门的事件管理。

表 5-1 城市运行事件管理矩阵

	交 通 事 件			食 品 安 全 事 件			X 事件
	数据 1	数据 2	数据 X	数据 1	数据 2	数据 X	数据 X
公安系统	I						
交通系统	O						

(续表)

	交 通 事 件			食 品 安 全 事 件			X 事件
	数据 1	数据 2	数据 X	数据 1	数据 2	数据 X	数据 X
卫生系统							
安监系统							
食药监系统							
社保系统							
民政系统							
教育系统							
环保系统							

注：“I”代表该系统需要获取该数据；“O”代表该系统输出该数据。

5) 接口管理

建设智慧城市,多系统整合是基础。为了更有效地管理多个系统接口,应该建立一套城市级的接口控制文档(Interface Control Document,ICD),该接口文档对应每个参与集成的各智慧应用,基于一定的标准规范,明确接口内容和对接方式。

一体化公共信息服务平台为领域系统提供的对接方式包括:

(1) WEB SERVICE。对行业各应用系统的接口的服务化封装,一体化平台提供服务注册、服务目录管理和服务质量管理等服务管理功能。

(2) 文件接口。对于批量大数据交互,通过约定数据文件格式,提供文件接口,实现数据文件的压缩、可靠传递,必要时提供文件加密、解密。

(3) 消息接口等。对于大量的实时信息尤其是事件信息采用消息(Message)方式。消息接口提供订阅/发布和广播等多种策略,并可以针对不同的接入系统进行消息的配置管理,对于订阅和发布消息类别,消息所带数据项、消息时效性等进行规范管理。

5.2.3 大数据平台服务

智慧城市应该是一个开放、互联的服务生态系统。市民、政府和社会各方都可以很方便地获取公开的数据,开发应用。为了简化各行业获取和应用大数据的方式,一体化平台应该基于云计算方式提供大数据应用引擎,也称为大数据平台服务,包括数据及服务(DaaS)和云平台服务(PaaS)。数据服务主要提供大数据检索、交换、订阅/发布、通知等信息服务,云平台服务侧重在基于云计算的计算资源管理,包括平台的高可用性和可伸缩性组件。

一体化平台处理的数据不仅量大,实时性高,访问量也非常巨大。基于云计算架构,实

现公共信息服务的虚拟化能提高平台的适应性。由于汇集了城市各行业需要共享的基础信息和城市运行的关键动态信息以及统计分析、模型等决策支持信息,这些信息相对于单个的行业系统而言将是海量的。一体化平台不仅要实现这些海量数据的采集、整合、存储、调阅等服务,并要保证服务质量。一体化公共信息服务平台将充分应用云计算技术,实现数据存储、数据管理和数据调用的虚拟化,将大数据按照数据类别、用户类别、数据活跃度等进行分解、并发和整合。

(1) 一体化平台的高可用性设计。一体化平台是智慧城市的信息枢纽,公共应用的支撑平台,其高可用性之重要毋庸置疑。一体化平台的高可用性技术实现是基于分解和冗余,即各项服务的分解、组件化封装,并通过划分主服务、备份服务和应急服务来实现 N 版本冗余。冗余是同一服务的多实例并行运行,并通过服务状态监管程序实现实时故障检测和故障点接管,以防止单点故障,提升系统可靠性。 N 版本($N \geq 2$)设计目的在于避免软件同源故障,例如一条异常消息可能会引起服务程序的崩溃,如果主服务和备份服务系统都是相同版本,则会同时出错。当然,从成本角度出发,主服务部件和备份服务部件可以是同版本的,运行在不同的物理环境上,但应急部件应该与主服务部分是不同的版本,包括同样功能的不同代码实现,以及功能上的简化。

(2) 一体化平台的可伸缩性设计。有两种情况对一体化公共信息服务平台提出可伸缩性设计的要求,即城市规模差异大和城市化高速推进。中国有超过 650 个城市,城市人口规模、管理模式各不相同,既有人口 2 000 万以上的特大城市,又有人口几百万的中大城市和人口几十万的小城市,其数据、应用规模也不尽相同;即使对于一个城市而言,城市规模也在飞速发展。一体化平台需要具备可伸缩性来应对城市规模差异,以及未来城市规模高速发展。基于云计算技术,可以实现计算资源的动态调配,按需申请,以保证一体化公共信息服务平台的可伸缩性。

5.3 大数据开放服务平台

大数据要发挥巨大价值,开放是关键。2013 年 5 月 9 日,美国总统奥巴马签署了一项行政命令,默认政府信息开放与机器可读。美国认为,数据就是重要的国家资产,开放数据能促进就业。例如开发气象数据和全球卫星定位(GPS)数据,极大地促进了产业的发展。

美国在政府数据开放方面一定有了实际的成果,Data.gov 已开通,Data.gov 按原始数据、地理数据和数据工具三个门类组织开放的数据。美国商务部 CIO Simon Szykman 在一次政府 IT 会议上指出“我们讨论的重心完全围绕着大数据,不过政府实现预期目标的真正障碍不仅在于收集数据,而是如何将数据转化为切实可用的信息产品以及开发知识”。美国能源部提供了一系列新应用程序接口(APIs),让软件开发者能获取诸如一个太阳能资

源探测器工具、每英里机动车耗油量预测以及实用速度数据库等。最近美国能源部启动了美国能源数据挑战计划,以利用能源数据解决美国面临的一些最紧迫的能源压力的伟大创意。美国农业部的特色数据集有美国农业部农贸市场目录和 API、经济研究服务可视化与数据集,以及一个国家农业统计服务的动态 API。现在公众也可以获得有关全球植物基因库的重要研究信息以及有关美国作物状况的新的卫星数据和绘图工具。

智慧城市建立统一的大数据开放服务平台,有助于打破信息孤岛,并实现数据授权,有助于发展数据产业。智慧城市大数据开放服务的基本思路如下:

- 政府数据开放、机器可读
- 应用云计算等新模式
- 鼓励社会化的数据再开发
- 城市一体化而不是部门孤岛

大数据开放服务平台要实现的基本目标如图 5-12 所示:

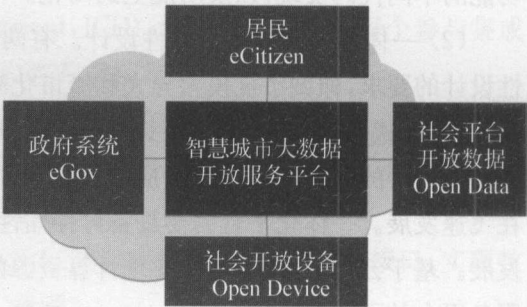
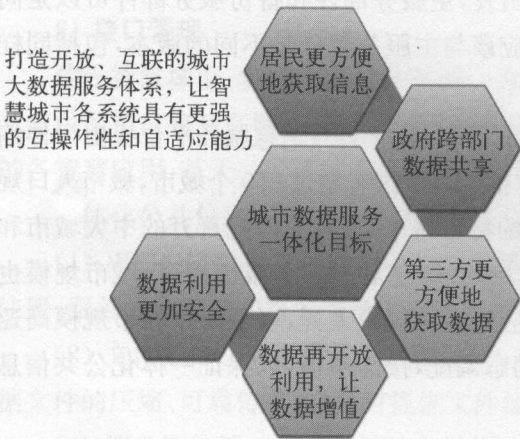


图 5-13 大数据服务生态环境

智慧城市大数据开发服务平台不仅仅是把政府各部门掌握的大数据进行开发共享,也不仅仅是为政府管理服务,该平台将连接居民、政府、社会开放设备和平台,形成一个大数据服务生态环境,如图 5-13 所示。

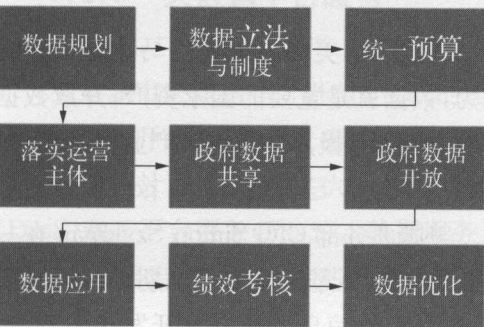


图 5-14 服务平台实现路径

当然,由于目前部门分散管理的模式,推进城市数据共享,以一体化思路做城市大数据服务并不容易。大数据开放服务难免会涉及部门利益,各部门在完成本部门业务操作和数据管理的基础上,很难有动力去推动跨部门的数据共享与应用协同。因此,大数据开放服务需要有顶层设计,这个顶层设计是在智慧城市总体规划指导下,推进数据共享立法,并把握住信息

化预算,明确责任主体,明确考核机制,并在运营过程中不断优化。其路径图如图 5-14 所示。

5.4 大数据决策支持平台

大数据能提升智慧城市决策支持的能力,形成更智能的趋势分析。中国工程院院士邬贺铨曾举了一个生动的例子:通过城市食盐的销量来计算出人口的数量,分析城市流动人口变化趋势。2012 年,广东为了比较准确地知道还有多少外来务工人员留在广东,广东东莞统计了 2012 年 1~6 月当地食盐销售量。因为食盐消费是刚性的,经济形势好的时候人们不会多吃盐,经济形势不好的时候也不会少吃盐,食盐的销量与人口数量是直接相关的。统计结果表明,2012 年 1~6 月,东莞的食盐消费量确实减少了,减少的比例与东莞的电信运营商根据 2012 年上半年活跃移动用户数的变化所推断出的数据几乎相当。由此推算留粤外来务工人员数量的变化是相当准确的。

可以预见,通过深入地分析和挖掘,大数据可以为智慧城市的很多领域服务,包括交通运行、医疗卫生、社会管理、农业、商业、金融以及宏观经济。当然,大数据要能为决策支持服务,仅有量大还不行,还需建立一套大数据技术与业务深度融合的模式,简化各领域应用大数据进行分析决策的难度,把大数据的决策分析作为一项服务来提供。大数据价值模型如图 5-15 所示。

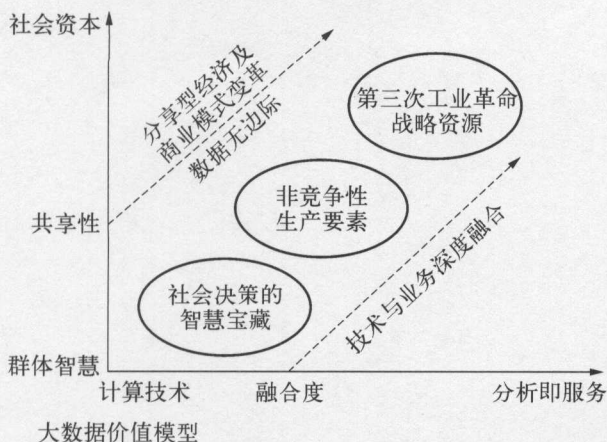


图 5-15 大数据价值模型

(图片来源:赛迪网-中国计算机报)

大数据的分析即服务,提供大数据分析的工具引擎包括多源数据关联分析、数据建模、数据仿真、数据报告等共性功能。

◇参 考 文 献◇

- [1] IDC, 6th EMC-sponsored Digital Universe Study. <http://www.emc.com/leadership/programs/digital-universe.htm>.
- [2] 中华人民共和国工业和信息化部. 2014年1月份通信业经济运行情况. 2014.
- [3] 维克托·迈尔-舍恩伯格, 肯尼思·库克耶. 大数据时代——生活、工作与思维的大变革. 盛杨燕, 周涛, 译. 杭州: 浙江人民出版社, 2013.



图 5-14 服务平台技术架构

第6章

大数据在智慧城市 中的行业应用

本章主要内容包括：智慧城市大数据应用的民生领域、监管领域、政府服务领域和基础设施领域的大数据分析、典型应用场景和应用趋势。

6.1 智慧城市大数据应用的民生领域

什么是民生？民生一词最早出现在《左传·宣公十二年》中，“民生在勤，勤则不匮”，这里的“民”就是百姓之意；《辞海》中对于民生的解释是“人民的生计”。孙中山先生也曾在《民生主义》中指出“民生就是人民的生活，社会的生存，国民的生计，群众的生命。”

智慧城市所关注的民生不仅限于最基本的“生存”和“生计”，更要以全面、深入、主动、个性的民生服务达到更高的层面，让城市“生活”更便利、更优雅、更精彩，实现“生命”的价值。而要达成这一目标，可以用物联网技术采集数据、用云计算技术处理数据、用移动互联网技术访问数据，还必须依靠大数据带来的智慧。

6.1.1 领域大数据分析

民生是以“人”为中心的。智慧城市的民生大数据自然围绕生活在城市中的人来展开。何人(Who)何时(When)何地(Where)有何需求(What)？就是服务提供者所需要回答的问题，而且不应该被动等待这样的询问，而是能够想在需求者前面，提供更主动、更体贴、更精准的服务。

如图6-1所示，达·芬奇著名的“标准人体”描述的是物理世界的人，如果将其投影到网络世界，就成为一位数字化的“虚拟市民”，两者有类似的属性构成，如：用于识别身份的头部、体现爱好和需要的躯体、维护社会关系和交往的手、留下行动痕迹的足。前两部分属性相对静态、稳定，后两部分属性则倾向于动态、可变。相异属性反映了不同的人的特征。那么，如果能够从匹配人的属性、满足人的需求出发来考察民生服务，就能做到针对性和个性化。

为了服务好虚拟市民，数据无疑是关键。许多人已经从购物网站提供的数据服务中尝到过甜头：当用户购买了一件商品时，网站会贴心地告知买这件商品的其他用户通常还买了什么，因为需求会有相似性，所以经常会对用户富有启发。虽然对于购物网站而言只是做了一个简单快速的数据统计而已，但对用户已经很有帮助（当然服务方也从中获益）。这就是数据的价值。

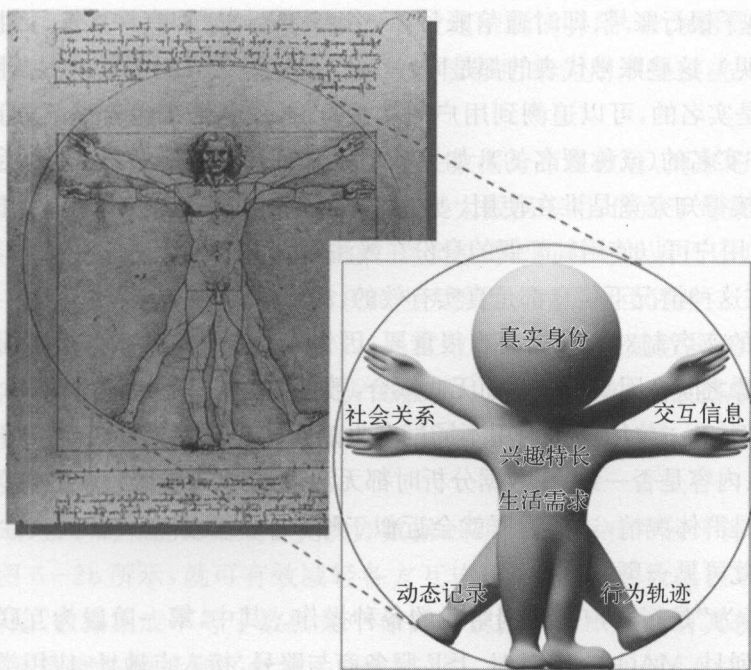


图 6-1 网络世界虚拟市民属性构成示意图

数字化民生服务与现实生活需求密切相关,许多服务更是贯穿线上、线下(Online to Offline, O2O),涵盖衣食住行、健康娱乐各个方面。在建设智慧城市的浪潮中,国内许多城市实际上已经走完了基础信息数字化这一步,市民档案、公共设施、服务资源等大多已形成较为完整的数据库,积累了庞大的数据量,为信息化民生服务的开展奠定了扎实的基础。只不过这些数据往往分散在不同的拥有者手中,没有被整合起来,也就形成了所谓的信息孤岛。

于是就会发生这样的情况:住房已经合法办理了转让手续,市民却还要专程到公用事业单位办理水电煤账户转移,否则倘若发生恶意欠费,个人信用负面记录可能还会算在前房主身上,因为住房交易信息没有被合理共享和利用。

事实上,当我们躺在家里沙发上点点键盘和鼠标就可以从互联网上找到各种信息,不必一趟趟跑图书馆辛苦翻阅书籍资料,就早已体会到了数字化的妙处和威力,据此可以大胆推想,一旦民生信息化服务得到完善,将使城市生活便利到何种程度!然而,民生领域既庞大又复杂,从来就不是个别的、局部的、单纯的事项,涉及政府职能部门、服务机构、公益组织、商业企业等诸多实体、不同层级,需要进行通盘的考虑、整体的规划、全局的设计。

1) 民生大数据的来源——数据获取

(1) 问题一:民生大数据从何而来?简而言之,民生大数据无非是关于人和行为的数据,换句话说,就是“谁干了什么”的记录,技术上可以记为{ID, EVENT}矢量形式,从属参数有时间(TIME)、位置(LOCATION)等。

网络上的“人”,即虚拟市民,体现为各种账号(Account),例如:电子邮件信箱账号、购

物网站账号、电子银行账号、即时通信账号,一个人可以同时拥有许多账号(拥有上百个账号的人并不罕见),这些账号代表的都是同一个人,是这个人在网上的映射。网络账号一般有两类:一类是实名的,可以追溯到用户的真实身份,或者至少包含一部分真实的身份信息;另一类是非实名的(或称匿名的),如免费电子邮件信箱,虽然使用者是活生生的人,但无法从账号直接得知究竟是谁在使用。此外,即便是实名账号,也需要具体区分其真实程度,例如:一位用户可以使用其配偶的身份在线注册购物网站账号,并一直用其配偶的银行卡账户支付,在这种情况下信息都是真实有效的,却并不能与“人”严格对位。

网络用户的实名制对大数据而言很重要,因为人的身份是联系各个方面数据的主键,否则各个数据源之间会因缺少关联而无法耦合,数据关系难以建立,这不是大数据,只是大量数据而已。例如,一位用户分别用不同网名发表了多个匿名网帖,则无论这些网帖是否在同一网站上、内容是否一致,在数据分析时都无法聚焦到同一个人身上,以至于不能确定是个体言论还是群体舆情,这类数据就会近似于噪声。大数据应具备真实性(Veracity)特征的重要性由此可见一斑。

网上的“行为”是指用户接入网络后的各种操作。其中,第一阶段为互联网接入,包括终端型号与序列号、MAC与IP地址、ISP服务商与账号、接入点地址、应用类型、访问时长等;第二阶段为网络应用,用户会根据需要访问不同网站、使用不同应用,操作可能包括注册登录、搜索内容、购买商品、递交申请、上传照片、沟通交流等。用户的操作都会在互联网上留下数字足迹,哪怕是点击了一篇文章链接、删除了一张相册照片,均可以被忠实记录下来。这些记录下来的数据就是回溯、统计、分析的基础,也是大数据的构成要素。

值得特别关注的是,近年来日益繁荣和普及的移动互联应用给用户上网行为带来了一些新的特点:由于用户使用智能终端保持随时在线,所记录的数据量可能呈数量级增加;又由于移动终端的位置会不断变化,增加了动态的位置数据,而这些数据恰恰是移动互联应用所特有的LBS(Location-Based Service,基于位置的服务)运行的基础。

除此以外,智慧城市中采集的人口数据、环境数据、交通数据、健康数据、监控数据、经济数据等也是民生大数据的重要组成部分。

综上所述,用户相关的民生大数据的采集是在用户参与下,由运营商、服务方分别采集完成的。

(2) 问题二:谁是民生大数据的拥有者? 比较容易回答的问题是:这些数据保存在哪里。其实从上文已经知道:除了一些“阅后即焚”之类的特殊网络应用外,用户相关的各种身份和行为数据都被保存在网络服务方的系统内,且往往没有期限。这些系统既是数据的生产(生成)者,又是数据的处理者和存储者。

进一步的问题是:这些数据究竟应该属于谁? 既然是用户的身份和行为,那么用户就是数据的主人,是数据的合法“拥有者”,但现实情况是,用户通常只能查阅到部分数据,很难掌握全部数据,而作为“保管者”的服务方则通过显性或隐性“征得用户同意”的方式保留和使用所有数据,拥有彻底的使用权,比主人更像主人。不幸的是,除了法律明文规定用户

数据不能买卖和泄露外,这些数据的使用并没有设定明确的界限,作为弱势一方的用户只能期待服务方遵纪守法并且有公德心,希望数据能够用来为用户提供进一步的、更好的服务。

(3) 问题三:如何获取民生大数据?民生相关数据分布在政府、商家的各个信息平台中,如果保持现状,虽不能绝对地说每家都不能开展大数据应用,但由于每家都缺少其他方面数据的支持,可能在数据量、多样性、更新率、真实性等不同指标上有所不足,难以达到理想的应用效果。因此,数据开放、共享是实现民生大数据应用的必由之路。

要获取他方数据,开放自身数据先行。通常数据提供方也是数据的需求方。如图6-2a所示,需求方可以按需获取其他提供方的数据,以实现本身的大数据应用目标。但是数据分散(分别)获取方式存在明显的缺陷,每方需要开发和维护过多的数据接口,管理工作量巨大,安全性也难以长效保障。

倘若有一家服务商有能力集成各提供方数据,采用统一标准和云计算技术为各个需求方服务,如图6-2b所示,就可有效减轻各方互连压力,是实现数据获取的合理方案。但需要说明的是,数据集成不等于数据集中存储。数据大集中虽然有其便利性,然而由此带来的问题是数据传输量和冗余量大、成本过高、数据利用率低、不适应高动态性数据,也会产生额外的安全隐患。可行的数据集成方案之一是集中式、分布式存储相结合,静态数据、高频数据、已读数据可集中存储,动态数据、更新数据则采用分布式存储并按需获取。

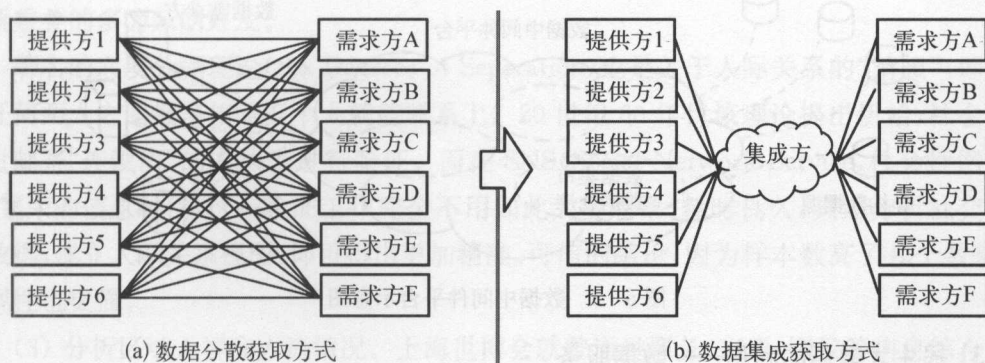


图 6-2 数据获取方式比较

数据服务(Data Service)是伴随大数据应用而产生的,核心思想是数据与网络服务的分离。传统的信息系统把数据和应用紧密绑定在一起,数据一般是从属于应用服务的“仆役”。而数据服务则赋予了数据更强的独立性,更重视数据的价值,使数据在系统中的地位上升到前所未有的高度。人们逐步意识到,数据不但可以为自身系统运行服务,也可以为第三方提供数据查询、数据中转、数据营销等服务,当然还可以运用大数据技术点石成金。

2) 民生大数据的汇聚——数据集市

教育、医疗、就医、住房、社保、分配、养老、居家、出行、安全、权益、助残、帮困……都是城市居民关系的切身问题，都是民生大事。民生大数据就是在民生信息化的基础上提升到更高的层面，不仅达到“有服务”的基本要求，而且能够让民生服务更便捷、更主动、更个性化、更有前瞻性。

在厘清了民生数据来源与获取相关问题后，后续的挑战是如何为民生大数据应用创造有利的发展环境。

在数据利用中最无法容忍的是“僵化”的数据供给——非规范化接口、结构一成不变、粗粒度、牵一发而动全身、拒绝按需应变、难以维护。尤其对于具有广泛来源、应用空间巨大的民生数据而言，理应在大数据应用发展初期就有意识地建立起效率高、适应能力强、扩展性好、标准化的数据中心。

如图 6-3 所示，可采用所谓数据中间件(Data Middleware)的技术架构，便于集成跨平台、异构化的数据，为数据需求者提供统一的服务。进一步地，在数据中间件平台上建立面向决策的专业化的数据集市(Data Mart)，逐步扩展为数据脉络更清晰的数据仓库(Data Houseware)，为大数据应用奠定扎实的技术基础。利用汇聚数据中间件平台可发布特定用户群体所需的数据，且无须受制于源系统的大量需求和操作性危机；支持访问非易变(nonvolatile)的业务信息；调和来自多个运行系统的数据；通过数据清洗从而提供净化的(cleansed)数据。

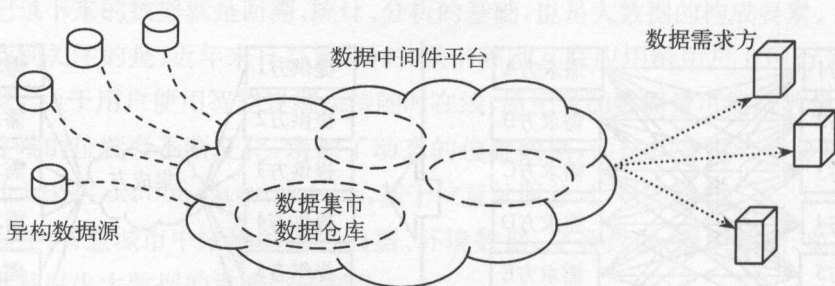


图 6-3 数据中间件平台示意图

3) 民生大数据的利用——数据服务

民生大数据可望在民生领域的舆情、预测、决策、调控等方面发挥巨大的作用。民生服务对象是人，那么不妨从一个与人(人群)的活动规律分析的例子来考察大数据的智能是从何而来的。

(1) 观察一个特定个体的活动。人的活动表现为地理位置的变化，但位置数据不限于 GPS 信息，而是来源多样：手机连接的基站、打电话、接入或靠近的 Wi-Fi 热点、蓝牙或 NFC 感应、刷交通卡或门禁卡、在 POS 或 ATM 机上刷银行卡、交通摄像头捕捉到的车牌号码、登机、看病、借书、公司或家庭电脑上网、看 IPTV 电视……这些所作所为均可以“暴露”人的行踪轨迹。假定每整点取得位置数据，以矢量{ HOUR, LOCATION}在地图上标

定,如图 6-4 所示,将 24 张标定地图进行图层叠加,则一天 24 小时的活动情况即可一目了然。如果积累一段时间的观察结果,可以判断出大部分晚间时分所处位置为“家庭”、大部分工作日白天所处位置为“公司”、上下班的时间、路径、路程、用时以及采用的交通工具。

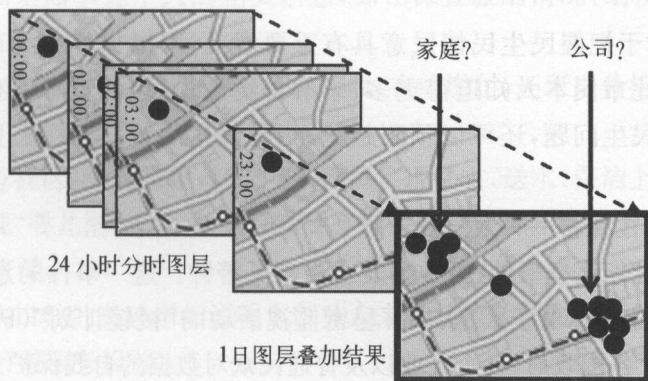


图 6-4 活动位置分析原理图

(2) 进一步梳理社会关系。对比不同个体的位置/位移图,结合其他数据,可以完善人与人之间的社会关系。例如:家庭位置基本重合的人可能是家人、邻居,可从通信联系及其频率加以区分;上班地点相似且存在通信联系的人则可判断为同事;虽然位置不重合但经常通信的或许是亲戚、朋友、同学、客户,如果工作日联系较多一般为客户,周末和假日发生位置重叠的多为亲朋好友。

著名的六度分隔理论(Six Degrees of Separation)也是关于人际关系的“猜想”,即地球上任何两人之间最多相隔五个人就能联系上。20 世纪 60 年代该理论提出伊始,社会学家通过邮寄“连锁信”的方法来进行验证。而在 SNS(Social Network Service,社会性网络服务)繁荣的信息时代,这一验证工作完全不用如此大费周章、耗时良久,只需分析社交网站的数据,建立人际关系视图,即可得出更加精准、可信的结论,因为样本数高了几个数量级,客观性也更强。

(3) 分析区域人群和人流情况。上海世博会就曾通过票务、交通、图像等信息综合分析后发布观展人数预警、预报,上海某商圈采用移动通信基站数据分析人流构成、流向等情况,都是向数据要信息、要智慧的案例。

类似地,通过大数据分析可以获得一个城市或一个区域人流变化的潮汐图、人群聚集的突发情况、人口迁移趋势等诸多信息,有利于民生问题的服务规划、资源部署、防控调度、政策制定和管理决策。

在民生领域的其他方面,大数据应用具有更大的想像空间。例如从感冒药销量的上升数据预测流行病的发生,可及时发出预警。再例如,近年业界有一个热议的有趣话题:用户需求调查还有必要吗?众所周知,服务行业最希望了解的是用户需要什么、哪些服务产品

能够吸引用户,因此唯恐忽略了用户的感受。那么为何有如此一问?理由之一是从网站数据分析就可揭示用户的喜好和趋向,没有必要发放传统的调查问卷。虽然这一观点过于绝对化,但是的确说明了大数据可以发掘更内在的、更准确的东西。比如一个人可能在调查问卷上回答自己月薪4 000元以下,实际上从系统数据上反映出其收入远不止这些,而具有更强的购买力。

大数据技术对于把握民生民情民意具有重要意义,可以提炼广大市民的需求,使民生服务领先一步,比市民本人知道得更多、了解得更细、发现得更早,可以分析社会现象本质,揭露潜在的民生问题,还可以预测趋势,指导民生决策,使立法、监管先行,防患于未然。

4) 民生大数据的保护——数据安全

2013年发生的“斯诺登事件”给数据安全敲响了警钟。这一事件的意义不仅在于曝光了美国政府广泛搜集全球电子数据、从事秘密监视活动的“棱镜计划”(PRISM),更重要的是其启示性,唤起了各国政府、各个企业以及普通民众对数据的自我保护意识。

由于民生数据涉及所有城市居民的隐私信息和政府机构的重要信息,安全保障无疑是重中之重。如果采用数据中心等集中存储方式,更要对网络和信息系统严防死守,杜绝“一锅端”式的信息灾难。

其实从大数据技术的角度来看,个人敏感数据泄露的威胁相对而言较小,因为民生大数据主要关注的是群体问题,单一个人数据在其中相当于沧海一粟,其份额和作用是非常微小的。当然,既然包含个人数据在内,就不允许发生哪怕是个别的外泄,这是民生大数据的“红线”。

从信息系统的角度考察,应当从数据传输、处理、存储的各个环节进行保护,切实防范数据窃取、盗用、伪造、篡改、破坏等安全攻击。

6.1.2 典型应用场景

2010年上海世博会有一句响亮的口号“城市让生活更美好”(Better City, Better Life),道出了生活在繁华但拥挤的城市中的人们的愿景。那么,城市是否能让市民的生活更美好?城市如何使市民生活得更美好?这里通过两个智慧城市中民生大数据应用的场景来回答这个问题。

1) 场景一:科学的人口管理,贴心的市民服务

户籍制度只能保证对城市常住人口、流动人口信息的准确记录,得到的人口数据库是相对静态的数据集合,仅仅在统计上有一些价值,与百姓生活更是毫无关联。但如果引入时间维度,情况就会大不一样。

对比历年来某个区域人口数据库,第一可以看到该地区人口数变化的情况和趋势,第二可以进一步观察到常住人口、流动人口相对变化的状况以及出生率、老龄化的状况

和趋势,第三可以再进一步揭示人户分离、婚姻特点、区域对人才的吸引力等深层次问题。

在此基础上,结合服务资源,通过大数据分析,就可发掘出先行一步的、聪明的民生服务举措:

(1) 市民可以全面掌握个人的相关信息,如社保金缴纳情况、纳税情况、个人信用情况等。

(2) 家庭的水电煤账单不但可以电子化接收、在线缴付,还可以分析并给出节能省钱的指导。

(3) 日常服务社区化、网络化,轻松点击就可实现洗衣、送水、保洁上门服务。

(4) 一旦出现“婴儿潮”现象,3年后就要有针对性地增加学前教育资源配置,6年后扩展到中小学教学资源。

(5) 工作日人口数的急速增加说明大部分人来自其他区域,并非居住在本区域,则最需要开发“白领午餐”等服务设施。

(6) 如果高学历人口加速外迁,说明本地区工作、生活环境缺乏吸引力,应推出相关吸引人才的政策、加强招商引资工作。

(7) 外来流动人口大量涌入时需着力配套公租房、廉租房设施,同时加强群租监管力度。

(8) 违章停车罚单数量的上升只能说明停车位紧缺或位置不合理,需发掘停车资源、规划停车场。

(9) 老龄人口尤其是孤寡老人数量应与养老机构资源成正比,并应配套相关的活动中心、社区餐厅、老年大学、引进家政服务人员、扩展助老服务志愿者队伍。

2) 场景二:智能交通,优雅出行

出行是几乎每个人每天都有的经历,是上班或上学的刚性需要。一次轻松的出行可以让人心情愉悦、效率倍增;而一次艰难的出行则会让人沮丧,并带来一系列的资源、金钱和时间的浪费。出行所遭遇的困难很多时候都是因为交通状况的不可预知,导致盲目性,没人乐意驾车陷入拥堵的长龙,也没人自愿加入摩肩接踵的人堆(狂欢节除外)。而预警、预测正是大数据的强项。物联网、大数据技术支撑下的城市智能交通将可以为市民的每一次出行保驾护航。

如果选择市内公共交通出行,只要告诉智能手机目的地,马上就能规划出轨道交通优先的换乘线路,下一班车的到站时间一目了然,从容出发,一路刷手机内置的交通卡即可乘车,到站前手机会自动提醒,并指出下一段公共交通站点的位置。

如果驾车出行,同样只需要告诉手机(或汽车)目的地,上车后手机自动将目的地信息同步给车载智能系统,并自动通过车联网了解沿途交通状况,导航仪指出的线路已经避开了交通高峰拥堵点,是一条既快速又省钱的最佳路径。万一前方道路突发拥堵,车联网也会第一时间将消息传递给行驶中的车辆,以便在可能的情况下重新规划、绕道通行。即将

到达目的地时,汽车会指出最近有空位的停车场,得到驾驶员确认后,将直接导航进入该停车场已经自动预定好的停车位,届时停车费也会自动扣款。智能化汽车会自诊断车辆运行情况,假如发现油量(电量)不足,提醒驾驶员的同时立即获取并显示最近的加油站(充电站)位置、油价信息,点击后即可导航前往,加油完成自动扣费。

6.1.3 应用趋势分析

大数据技术之所以被形象地称为“数字炼金术”,是将数据比喻为矿藏,而从数据中提炼有用信息的过程就像获取宝贵的金子,当然还有另一层意思,大数据就像淘金一样存在一定风险,假如数据质量不合格或方法不合理,或许根本找不到金子。

美国心理学家马斯洛对人类的需要进行过系统的、独到的研究,将其归纳为具有5个层次的“金字塔”,被学术界称为“需要层次理论”,从基本需求到最高需求依次为:生理、安全、社会、尊重、自我实现。民生大数据是为市民的需要服务的,也可与“马斯洛需求金字塔”对应起来:

(1) 生理——基本的民生需求,基本生存条件的保障,如衣食住行、义务教育和就业。

(2) 安全——让市民生活衣食无忧、有安全感,特别是病有所医、老有所依,守法公民得到法律的充分保护。

(3) 社会——安逸、和谐的小康生活,有稳定、融洽的社会关系,政民互动顺畅。

(4) 尊重——有一定的事业成就和社会地位,生活条件优越,渴望得到尊重,掌握一定的话语权。

(5) 自我实现——实现个人理想、抱负,个人能力发挥到最大程度,达到自我实现(或自我超越)境界。

民生工程必然要符合公平、公正性,所以必须能够满足各个层次的需要,照顾到所有人(至少是大多数人)的利益。民生金字塔的底层越厚实、社会基础就越扎实、上层建筑就越稳固、发展空间就越巨大。

基于数据的学术研究和技术应用被称为数据科学,则面向民生领域的大数据应用就是民生服务的科学化、智能化。数据来源于现实,但数据服务应当超越现实,推动社会进步和生活质量的提高。

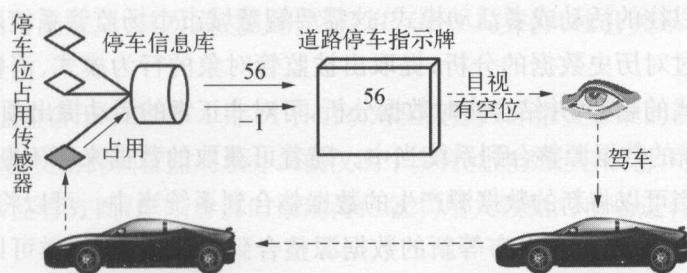
自人类社会进入信息时代以来,数据服务于社会总体上可划分为3个台阶:第一个台阶是数据的统计分析,基本上就是现实世界状况的直接反映,例如12岁的女孩有多少名、初中有几所;目前所处的第二个台阶是数据的分析挖掘,可以提取出潜在的、不明显的信息,例如从大量的信用卡消费记录中了解高收入人群最近喜欢在哪里购物;第三个台阶式是未来的大数据,可化未知为已知、化复杂为简单、化腐朽为神奇。

民生领域的信息化进程同样经历3个发展阶段:

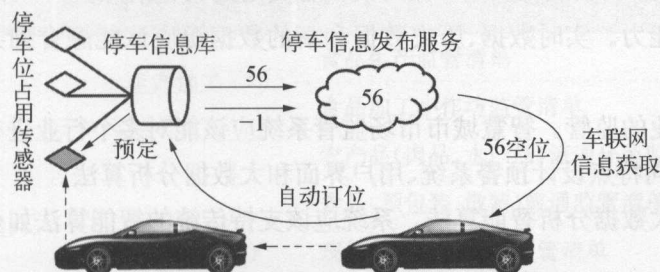
(1) 2010 年之前的初级阶段,各个政府部门、各个民生相关行业纷纷建设和上马信息处理系统,实现了资源的数字化、服务的信息化,例如内部办公和管理系统、审批系统、服务网站。这个阶段的特点是信息系统各自为政、相互割裂,数据分散在各自的系统中,条块间很少实现共享、互联。例如部门间互不了解对特定对象的帮困情况、婚姻状况不能跨省交换。这一阶段暴露出来的问题成为后期攻坚克难的重点目标。

(2) 自智慧城市概念的提出和建设的启动,进入目前的中级阶段。信息系统经过多年的运行积累了大量的数据,数据的重要性被意识到并被提升到相当的高度,在政府的主导下开始逐步实现数据的集成,如城市数据中心、公共信息服务平台等。数据的开放也被提到议事日程上来,从而不限于在政府部门间共享数据资源,而且对个人(企业)用户更加透明,可以预计不久的将来必然会对第三方应用服务全面开放。这个阶段中,数据服务的多样性也随着移动互联应用的普及而得到加强,例如原来实时交通路况信息只能通过交通台广播或上网站查询,现在可以直接推送到道路上的显示屏、智能手机、车载导航仪、商用地图等,但仍然停留在目视阶段,说明“智能服务”还离不开人的介入,如图 6-5a 所示。

未来的高级阶段将以大数据技术和应用为标志,智慧城市变得真正能够善解人意,信息服务环路中不需要人的干预,如图 6-5b 所示,智能化停车诱导系统与物联网车辆一起构成信息服务闭环,而达成理想目标的关键正是智慧城市的民生大数据服务。



(a) 目视式停车指示系统(开环)



(b) 智能化停车指示系统(闭环)

图 6-5 停车诱导系统比较

6.2 智慧城市大数据应用的市场监管领域

6.2.1 领域大数据分析

市场监管是指市场监管主体对市场活动主体及其行为进行限制、约束等直接干预活动的总和。市场监管的目的是维护市场运行的安全稳定、保护市场主体的合法权益,以及促进市场的开放和发展。不断变化的监管环境、不断增长的数据量、处理数据的复杂度不断增加,以及市场监管者的角色不断发生变化,这些都对当前市场监管系统技术提出了新的要求和挑战。基于大数据的下一代市场监管系统应该满足以下要求:

(1) 实时分析(Real-Time Analytics)。市场监管系统应该可以实时分析瞬息变化的市场数据,及时作出反应。在金融市场中,需要实时监控交易行为,准确定位可疑的内部交易和老鼠仓等违法交易并及时作出处理。在对食品安全的监管中,需要对食品安全事故及时作出反应,避免食品安全事故中的受害人群迅速扩大。因此实时计算和复杂事件分析处理能力是智慧城市市场监管系统的关键技术能力。

(2) 历史大数据的分析能力。市场监管系统必须在特定情境中进行事件分析。可疑的活动可能是不同以往的活动或者活动模式,这需要智慧城市市场监管系统对以往活动进行分析和处理。通过对历史数据的分析,提取出被监管对象的行为模式、消费者的行为模式以及他们行为模式的趋势,并结合实时数据分析,可对非正常的活动提出预警。

(3) 能够把新的数据源整合到系统当中。随着可获取的数据来源不断增多,智慧城市市场监管系统应当可以将新的数据源产生的数据整合到系统当中。可以将社交媒体(如微博)、新闻媒体报道、公司信息发 布等新的数据源整合到系统当中,这样可以提高系统对非正常事件进行预警的准确度。例如发现某公司对市场披露新的信息之前,该公司的股票交易有不正常的活动,那么系统将产生一条可疑内幕交易的警告信息。

(4) 数据融合能力。实时数据、历史数据和新的数据源应该能融合到共同的数据处理平台。

(5) 支持多行业的监管。智慧城市市场监管系统应该能对多个行业进行监管。这需要根据各个行业的不同特点设计预警系统、用户界面和大数据分析算法。

(6) 支持多种大数据分析智能算法。系统应该支持传统的智能算法如分类、聚类、文本挖掘等。

(7) 大数据分析智能算法应该具有可适应性。由于社会、文化和经济的发展,监管的要求也在不断变化。智能算法应该能适应监管要求的变化。

(8) 系统应该允许引入新的大数据分析智能算法。这也是社会环境以及监管要求的变

化对系统提出的要求。

(9) 灵活的流程管理和案件管理。市场监管者在处理案件时可以根据实际情况灵活应用不同的处理流程,在管理案件过程中也可以灵活对案件处理人的分配、案件处理进度的标注、案件处理过程当中的资源调度等进行处理。

(10) 信息系统基础架构兼顾大数据的处理要求和监管的业务要求。

2013年7月,原博时基金经理马乐老鼠仓案被控制,后被查出任职期间利用内幕消息炒作76只股票,涉及金额10亿元,非法获利1883万元。这是迄今为止发现的最大规模的老鼠仓。马乐案发便源于交易所的大数据监管。目前上海证券交易所(简称上交所)和深圳证券交易所(简称深交所)各有一套证券交易监控系统,主要就是分析各种看似杂乱而无关的分散信息,利用各种算法将数据进行自动关联、聚类、分类和重排,寻找各种违规交易的蛛丝马迹。

公开资料显示,上交所和深交所的证券交易监控系统,系统集成了交易、登记、结算数据和上市公司、证券公司等相关信息。上交所异动指标分为4大类72项,敏感信息分为3级,共11大类154项;深交所建立了9大报警指标体系,合计204个具体项目,其中包括典型内幕交易指标7个,市场操纵指标17个,价量异常指标15个。交易所针对老鼠仓等交易行为还建立了专项核查和定期报告制度。

交易所所有两套分析系统,分别针对网络上可以公开获得的信息和交易所内部各账户的交易信息,这些都属于大数据。大数据捕鼠大致分三步:首先是通过网络信息和交易数据的分析挖掘出可疑账户,其次通过分析交易IP、开户人身份、社会关系等进一步确认;最后进入调查阶段。

马乐案中,证监会于2013年4月11日启动对相关账户异常交易案的初查工作,2013年6月21日立案稽查,交由深圳证监局承办。仅仅两个多月就能锁定老鼠仓,可见大数据的威力。

在食品药品监管方面,广东省佛山市南海区就利用大数据对市场进行监管,见表6-1。

表6-1 南海区监管清单

监 管 分 类	监 管 环 节	清 单 名 称
食品(含保健品)	生产加工	食品生产监管清单
		食品加工小作坊监管清单
		农产品(肉品、水产品)流通环节监管清单
		食品(预包装、散装)流通监管清单
	流通(现场制售)	现场制售食品流通监管清单
		酒类流通监管清单
		保健食品经营监管清单
	消 费	餐饮服务监管清单

(续表)

监 管 分 类	监 管 环 节	清 单 名 称
药 品	流 通	药品(零售药店)监管清单
		药品(医疗机构)监管清单
化 妆 品	流 通	化妆品经营监管清单
医 疗 器 械	生 产	医疗器械(生产)监管清单
		医疗器械(经营)监管清单
	流 通	医疗器械(体外诊断试剂)经营监管清单
		医疗机构使用医疗器械监管清单

南海区发布了一份监管清单,范围涵盖南海食品药品监督管理局所有的职责领域,包括食品(含保健品)的生产加工、流通(现场制售)、消费、药品流通、化妆品流通以及医疗器械的生产、流通四大行业、三大环节,具体分为 15 大模块。

在建的食品药品智能监管平台,可以根据录入的经营主体信息,与 15 个模块比对、结合,动态生成各个经营主体独有的监管清单。今后执法人员完全可以对照平台对每家企业“量身定制”的清单,一一进行监管,这将改变过去执法随意性强的问题,使执法更公平更有效。伴随着监管内容标准化、公开化,市民参与度提高,以及举报反映限时办结等情况,将进一步倒逼食品药品监督管理流程再造,提高监管效率。

6.2.2 典型应用场景

下面分析市场监管中的一个典型应用:通过对消保委的消费者投诉数据进行文本挖掘,得到快递业存在的问题。

随着快递行业的飞速发展,快递在人民日常生活中的作用也越来越大,同时消费者对于快递行业的投诉仍时有发生。这类投诉中包含了很多有价值的信息,这些信息对于邮政监管和企业建设都有很大的作用,所以对这类超短文本进行快速的分类是很有必要的。与以往的文本挖掘相比,消费者投诉所记录的描述信息通常在 50 字以内,因而文本的词汇量很小而且关键词汇出现的频率很低,同时还存在着不规范和不完整的特点。传统针对大篇幅文章的分类方法因而在此不是很适用,当然在此之前有很多专家学者都对文本挖掘抑或是对短文本挖掘都有很多的研究^[1,2]。孟宪军等^[3]利用互联网信息来进行文本聚类与检索技术研究并取得了良好的效果,邓莎莎等^[4]运用支持决策机制的方法来研究文本分类,并且国内外学者提出了诸如贝叶斯网络、主题模型、决策树、文本相似度等聚类方式,但这些方式方法在长文本中的精度较高,面对超短文本这些方法的效果还是不够理想。

因此下面通过对消费者投诉上万条超短文本的研究,通过研究投诉词汇间的相关性来

建立相似词汇文本的方式,运用二叉树分类来对消费者投诉进行快速的归类。这在一定程度上提升了投诉数据库的分类效率,同时也为今后扩展到对其他行业投诉超短文本的归类有了一定的研究方法思路。

1) 快递行业发展概况

随着我国经济发展水平的快速提高,快递行业近年来受电子商务发展的影响,也进入了飞速发展的时期。国内快递行业也从以前的中国邮政一家独大到目前的百家争鸣,快递行业的发展推动了互联网经济,促进了劳动力就业,而快递行业作为最基础的服务行业也受到了人们的关注。快递业在国民经济中的比重和地位快速提升,在拉动内需、服务流通、扩大就业和改善民生等方面的作用进一步凸显。

上海市快递行业 2012 年实现业务收入为 182.82 亿元,业务量为 59 906.9 万件。2013 年实现快递业务收入 257.6 亿元,同比增长 40.9%,快递业务量 95 012.4 万件,同比增长 58.6%。与此同时,2012 年上海市交通运输、仓储和邮政业总产值为 895.31 亿元,2013 年为 935.06 亿元,因此上海市快递行业总收入占到了交通运输、仓储和邮政业总产值的 20% 以上。

中国的快递行业从 1980 年 EMS 成立开始起步,到 1993 年民营企业可以从事快递业务才结束了 EMS 一统天下的格局,2006 年外资企业可以在中国成立全资子公司,同时电子商务行业的快速发展使得中国的快递行业进入了飞跃发展的阶段。在 2000 年以前,中国的快递企业基本从事的是进出口业务公司的快递服务,客户主要集中在船运公司、制造商和贸易商。2000 年以后,诸多民营快递企业的不断设立,使得快递费用不断降低,与此同时互联网快速普及,电子商务飞速发展,进而带动了快递行业业务规模和收入的爆炸式增长^[5]。

2003 年以前快递行业 80%~90% 的业务量由中国邮政完成,而根据国家邮政局统计数据显示,截至 2013 年年底具有经营许可证的快递企业已经超过 9 000 家,其中 79% 左右的市场份额已经被 6 家过百亿的快递企业占据。2013 年我国快递日均业务量突破 3 000 万件,高峰期超 6 500 万件。全年快递业务量将达到 92 亿件左右,仅次于美国,位居世界第二。

2012 年,淘宝和天猫平台交付的邮件包裹数量占到中国所有包裹总量的 60%,日均包裹量为 1 200 万个,淘宝日均包裹量占中国快递业 60% 以上。就淘宝商城的统计数据来看,目前淘宝上大部分店铺分布在东部沿海区域,这些区域也产生了中国 30% 以上的快递业务量。

由于 C2C 电商(如淘宝等)主要以个体商户为代表,其收入和业务规模相对较小,自建快递服务的成本较高,因而使用规模以上快递企业的服务在所难免。另外快递成本逐步降低,使得快递成为网络零售的支撑。不同于 C2C 电商模式,B2C 电商(如京东、苏宁等)由于自建仓储中心,其快递更多地集中在同城业务,其使用规模以上快递企业的服务相比 C2C 电商相对较小。但中国幅员辽阔,单个企业难以覆盖所有的快递区域,因而跨地域快递和海外快递仍由专业的快递企业提供服务。

同时根据相关数据统计显示,GDP自然增长1%,快递收入增长将达到2.5%以上。快递行业的发展反映着经济结构的变化以及经济的增长速度,快递行业的需求还与企业数量、人口数量、国际化进程和互联网发展密切相关。

2) 文本挖掘研究概要

文本挖掘最早用于信息的检索,20世纪50~60年代,Luhn、Maron、Kuhns对其进行了研究。70年代,Jardine、Van Rijsbergen开始针对性地研究文档聚类和文档相似度,然后Salton开始用频率较高的动词来表示一个文档。随着20世纪80年代人工智能和机器学习的逐步兴起,Hayes和Weinstein逐步定义了现在所正式研究的文本分类^[6]。

对中文文本进行挖掘需要先对文本进行中文分词和建立语料库,在这个方面已经有很多的研究。语料库主要是存放语言材料的数据库,其中J. Thomas和G. N. Leech^[6]为语料库的主要奠基人和倡导者。目前主要的中文语料库有由北京大学计算语言学研究所2002年完成的北京大学语料库(俞士汶等),以及中国台湾中研院于20世纪90年代建立的汉语平衡语料库和汉语树库,另外还包括美国宾夕法尼亚大学开发的LDC中文树库和欧盟开发的LC-STAR语料库。这些语料库的建立极大地促进了中文文本分析的发展,对于中文本身的复杂性,很多研究都针对性地提出了相应的问题和解决办法。

由于每一个句子都是由词汇组成,因而机器理解自然语句还需要知道词汇的信息,因此词汇库在处理自然语言有着重要的作用。目前常用的词汇知识库有美国普林斯顿大学开发的WordNet、日本电子词典研究员开发的EDR、以汉语和英语词汇所代表的概念为描述对象的HowNet。

当对文本进行分词、提取系数矩阵等结构化处理后,还需要对文本进行聚类分析。常用的聚类分析方法包括朴素贝叶斯分类法、基于支持向量机的分类器、K-means法、神经网络法、决策树法、模糊分类法等。

上述这些分类方法用于长文本的分类处理,对于投诉这类超短文本的分类,其效果还是不理想,这与短文本的稀疏性和不规则性有很大的关系。同时随着微博、Twitter等网络社交短评论性语句的出现,不少学者开始研究如何将短文本进行归类处理。国外学者对于短文的研究主要集中在相似度计算和文档矩阵稀疏性的处理上,Mehran Sahami^[7]提出了基于Web语义核函数的进行相似度研究的方法,Papadimitriou^[8]和Thomas Hofmann^[9]提出了使用隐含主题来针对性地解决短文本稀疏性问题,不少学者也提出使用关键词共现或查询词拓展技术。

当然国内学者也对短文本分析做了很深入的研究,这类研究首先集中于对中文文本的分词上。由于中文不像英文那样每个单词之间存在着空格,而且中文语句间词汇的使用方式因语境而变,因而中文的文本分词也成了对中文进行文本分析的首要问题。李振星等人^[10]运用全二分法来快速地进行分词,湛燕等人^[11]研究基于中文文本的分词方法,费洪晓等^[12]利用词频来研究中文分词。

其次,在对短文本的研究上主要集中于对词汇的特征处理和分类的算法上。彭泽映等

人^[13]分析大规模短文本的聚类方法,张志飞等人^[14]研究基于 LDA 模型的短文本分类研究,孙建旺等人^[15]基于文本语义来对短文本进行分类,蔡月红等^[16]运用半监督法来对文本进行分类。这类方法对于短文本的分类确实有不错的效果,但是对于投诉这类小于 50 字的文本来说效果还是不够理想。

3) 上海市快递行业投诉超短文本分析方法

基于上海市消费者权益保护委员会所提供的消费者投诉数据库,利用 R 语言的文本挖掘技术,将消费者的投诉具体描述信息抽取形成语料库,然后对这些语料库进行停词处理和解析抽取后建立结构化的词汇文档矩阵。举例来说,抽取消费者投诉数据库中的两条投诉描述信息,其原始表达如下所示:

- (1) “去年 12 月 24 日从上海到杭州的快件,至今未收到。现投诉,要求赔偿。”
- (2) “12 月 28 日委托该公司托运一条裤子至杭州,由于至今未送到,交涉多次未果,现求助尽快解决。”

对以上两句投诉描述信息,利用 R 语言中的 Rwordseq 软件包进行中文分词后得到如下分词结果:

- (1) “去年”“12 月”“24 日”“从”“上海”“到”“杭州”“的”“快件”“至今”“未收到”“现”“投诉”“要求”“赔偿”。
- (2) “12 月”“28 日”“委托”“该”“公司”“托运”“一条”“裤子”“至”“杭州”“由于”“至今”“未送到”“交涉”“多次”“未果”“现”“求助”“尽快”“解决”。

得到以上两句分词结构后,利用 R 语言中的 tm 包中的 Corpus 函数对其处理形成相应的语料库,然后通过 tm 包中的 tm_map 函数移除语料库中多余的空白和数字等无用信息最后通过 TermDocumentMatrix 函数形成如表 6-2 所示的词汇-文档关系稀疏矩阵。

表 6-2 词汇-文档关系稀疏矩阵

中文分词词汇	描述语句 1	描述语句 2
多 次	0	1
交 涉	0	1
解 决	0	1
尽 快	0	1
快 件	1	0
赔 偿	1	0
投 诉	1	0
托 运	0	1

(续表)

中文分词词汇	描述语句 1	描述语句 2
未 果	0	1
要 求	1	0
一 条	0	1
至 今	1	1
未收到	1	0
未送到	0	1

在得到词汇-文档关系稀疏矩阵后,下一步需要对每一个投诉描述语句进行归类,以便快速地对大量的投诉数据进行分析。目前常见的分类主要是根据文本的特征来进行划分,主要的分类方法有贝叶斯网络分类法、神经网络法、K-means 法、模糊分类法、决策树法和 SVM 分类法等。

但并非上述的每一类方法都适用于短文本分类,贝叶斯网络分类法主要利用 $P(h|D) = P(h) * P(D|h) / P(D)$ 来求取 $P(h|D)$ 的大小,来判断观测数据 D 是属于哪一类聚类空间 $H(h_1, h_2, h_3 \text{ 等})$ 。由于 $P(D)$ 都是一样的,所以在比较 $P(h_1|D)$ 、 $P(h_2|D)$ 等的大小时只需要比较 $P(h) * P(D|h)$ 的大小,在每一种分词下每一个投诉描述性语句在除去分词符外总是正确的假设下, $P(D|h)$ 可以假定为 1。而 $P(h)$ 的大小取决于词汇-文档关系稀疏矩阵中出现的词频次数,因而词频次数大的归类的概率就高。

在短文本中关键词出现的词频次数往往很低,从而使得关键词的 $P(h)$ 概率很小,所以利用贝叶斯网络这类算法无法准确地对投诉这类短文本描述性语句按照关键词进行分类。

另外由于每年上海市针对快递行业的投诉数据量有上千条,每一个投诉描述性语句经过分词后会生成数十个词汇,因而每年的投诉都会形成上万个词汇。除了关键性的词汇如遗失、损坏等在单个投诉语句中只出现一次甚至不出现外,另外投诉性语句之间表达的意思虽然相同但是所用的词汇完全不同(如同样表达损坏,但语句中可能通过摔坏、破损来表达),因而导致短文本挖掘在归类时很难利用词频和相同词汇来准确地归类。所以在针对这类短文本分类时,还需要将词汇-文档关系稀疏矩阵中表达相同意思的词汇进行预处理。

在对词汇-文档关系稀疏矩阵预处理前需要预先确定划分类别,从而合理地找出表达这些类别的描述词汇。本章中通过对快递行业投诉的词频和词汇间相关性的处理分析,可以从 2013 年消费者投诉词汇聚类分析图(图 6-6)中初步得出,消费者主要对快递行业的延误、损坏、遗失、费用不合理、售后服务不满意。进一步通过对每年所有消费者投诉的描述性语句的统计词频大小分析,以及通过用搜狗输入法的词汇字典针对相关行业的词汇描

述,针对性地提取出了以下几类:将投诉语句中出现{遗失、丢失、找不到、丢件、下落不明、缺失、弄丢、搞丢、寄丢等}归类到“遗失”这一类中,将{损坏、破损、破碎、碎了、破了、损毁等}归类到“损坏”这一类中,将{延误、延迟、未收、尽快送、拖延、未送、延时、迟迟不送、没送到、没发货等}归类到“延误”这一类中,将{售后服务、送货上门、先验货后签收、态度恶劣、态度差等}归类到“售后服务”这一类中,将{费用不合理、包装费、拒付、重复收取、支付等}归类到“费用不合理”这一类中,剩余的归类到“其他”中。

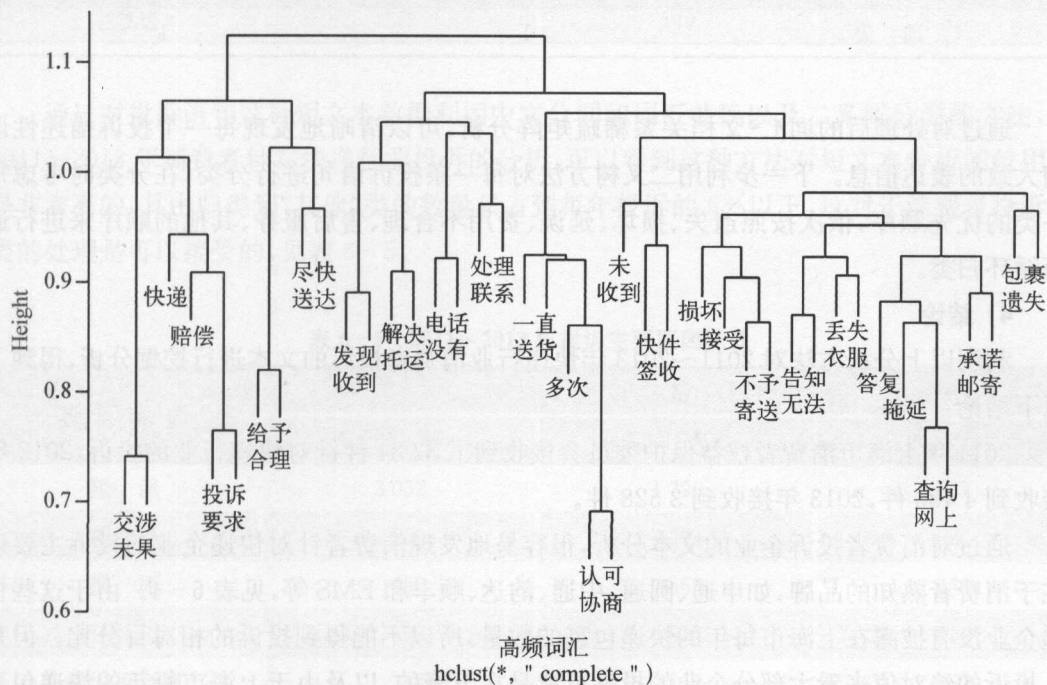


图 6-6 2013 年消费者投诉词汇聚类分析

由于消费者在针对快递的同一次投诉中会对几类问题进行描述,如同是出现破损和延时两类问题。但是为了不重复计算投诉,因而在归类时对归类顺序按照性质的严重程度做了相应的排序,如将“遗失”放在首位因为其性质更为严重,剩余的依次为“损坏”、“延误”、“售后服务”、“费用不合理”、“其他”。

利用以上思路,在算法中利用遍历循环将词汇-文档关系稀疏矩阵的每一列(即每一条)投诉语句进行处理。首先通过分析将对归类作用不大的分词词汇利用 removeWords 函数通过预先建立好的停用词表将它们去除,另外根据上述分析分别建立{遗失、损坏、延误、售后服务、费用不合理}归类词汇表,再利用遍历循环将关键的描述性词汇更改为相应的类别描述词汇。所以通过停用词处理和更名处理后,先前的例子中一些无用的词汇如“一条”、“托运”等就从矩阵中删除掉了,而“未收到”和“未送到”就替换成为“遗失”储存在新的词汇-文档关系稀疏矩阵中。在经过停用词和词汇替换处理后,词汇-文档关系稀疏矩阵的相关性就进一步提高,同时减小了矩阵的存储量,见表 6-3。

表 6-3 处理后的词汇-文档关系稀疏矩阵

中文分词词汇	描述语句 1	描述语句 2
尽 快	0	1
解 决	0	1
赔 偿	1	0
至 今	1	1
遗 失	1	1

通过对处理后的词汇-文档关系稀疏矩阵分析,可以清晰地发现每一个投诉描述性语句大致的表达信息。下一步利用二叉树方法对每一条投诉语句进行分类,在分类时考虑每一类的优先顺序,依次按照遗失、损坏、延误、费用不合理、售后服务、其他的顺序来进行遍历循环归类。

4) 结论

利用以上分析方法对 2011~2013 年快递行业消费者投诉的文本进行挖掘分析,得到了以下结论:

2011 年上海市消费者权益保护委员会接收到了 5 984 件针对快递行业的投诉,2012 年接收到 4 483 件,2013 年接收到 3 528 件。

通过对消费者投诉企业的文本分析,很容易地发现消费者针对快递企业的投诉主要还在于消费者熟知的品牌,如申通、圆通、中通、韵达、顺丰和 EMS 等,见表 6-4。由于这些快递企业没有披露在上海市每年的快递包裹的数量,所以不能得到投诉的相对百分比。但是从投诉的绝对值来看大部分企业的投诉总量是在下降的,以及由于上海市每年的快递包裹数量是成倍增长的,所以可以大致地推算出每个企业的投诉相对百分比也是有所下降的。这都说明物流快递行业整体的服务质量是逐步向好的,这也与快递行业人员增加、网店建设和汽车、飞机等硬件设施投入不无关系。

表 6-4 2011~2013 年投诉量较高的快递企业统计

投诉企业名称	投 诉 量 (件)		
	2011 年	2012 年	2013 年
申 通	1 034	659	477
圆 通	935	632	426
韵 达	864	983	727
中 通	673	339	59
汇 通	8	329	239

(续表)

投诉企业名称	投 诉 量 (件)		
	2011 年	2012 年	2013 年
顺 丰	184	229	310
EMS	386	278	292
天 天	279	239	170
宅急送	277	149	83

通过对投诉语句这种短文本数据利用中文分词和词汇替换以及二叉树分类的方法,对2011~2013年消费者针对快递行业投诉的分析,可以看到这种方法对短文本分析的效用还是非常高的,其中归类到“其他”类的数量只占到每年投诉的5%以下,这对于消费者投诉分类的处理是可以接受的,见表6-5。

表 6-5 2011~2013 年投诉主要原因统计

投 诉 原 因	投 诉 量 (件)		
	2011 年	2012 年	2013 年
延 误	3 052	1 793	1 217
遗 失	1 259	1 234	920
损 坏	690	866	550
售后服务	591	333	598
费用不合理	312	169	157
其 他	80	88	86
总 计	5 984	4 483	3 528

另外,通过各项投诉的处理分析,可以清晰地发现延误问题在近几年的投诉绝对数量中是呈现明显下降态势的,这也验证了物流快递企业近几年加大了物品投递及时性的处理。不论是顺丰的自营模式还是申通的加盟模式,都加大了在汽车和飞机等物流设施的投入,同时也扩大了快递投送人员队伍,这些措施都使得快递的时效性得到了很大的改善。但与此同时可以发现,在物品的损坏、售后服务等方面,消费者投诉的绝对数量没有明显的下降,虽然这些投诉问题与每年快递企业投递的包裹数量百分比是逐年降低的,但这些问题在每年投诉总量的占比是不断提升的。这说明在快递企业不断扩大规模时,企业员工的服务态度以及对投递物品的保护措施都还有待提高。

采用同样的处理方法,通过对售后服务的分项处理,发现消费者对快递行业售后服务不满的最主要原因是快递员的服务态度,然后是不送货上门,见表6-6。在快递行业的时

效性得到改善后,快递员的服务质量还需要提升,这也是我国快递业在快速发展阶段急需解决的问题,只有更好的服务质量才能赢得更多消费者的信赖。

表 6-6 2011~2013 年售后服务投诉原因

投 诉 原 因	投 诉 量 (件)		
	2011 年	2012 年	2013 年
服务态度	332	212	400
不送货上门	180	93	173
不让验货	79	28	25
总 计	591	333	598

5) 展望及进一步工作

通过对快递行业投诉短文本的分析,介绍了一种对短文本数据的处理方法,并通过对关键词间的相关性建立合理的聚类词汇文档,进而通过二叉树的方式来对每一条投诉数据进行归类分析。下一步将对现有的投诉数据进行分析,并研究出更适合处理投诉短文本的更优方法,从而更精确地将此类方法推广到各行业投诉数据的处理上。

另外就目前的处理结果来看,延误和遗失是快递行业主要的投诉原因,同时在上海市中快递行业的龙头企业“四通一达”和顺丰仍然是消费者的主要投诉对象。但根据国家统计局显示的国际快递数据,FED、UPS、DHL 投诉总量却相对较低。这些数据表明国内的快递企业与国外成熟的快递企业相比,在服务质量上仍有较大的提升空间。

6.2.3 应用趋势分析

未来的智慧城市市场监管有以下趋势:

- (1) 整合不同来源的数据进行市场监管,整合原有的监管机构,使之利用同一平台对市场进行监管,最后从单项监管向全面监管转变。
- (2) 对信息系统架构进行优化,使之具有很强的扩展性,能并行处理多项任务,有强大的计算能力。在此基础上以更快更灵活的方式进行数据挖掘,获得知识。
- (3) 监管流程更加公开,通过智能监管平台实时发布数据,使消费者即时得到市场监管相关信息,也使研究者拿到数据进行深入研究以便进一步改善大数据分析的智能算法,最后实现从封闭式的监管向开放式的监管转化。
- (4) 充分利用物联网的感应器和电子标签提供的信息,对商品尤其是食品药品进行监管。在监管的同时也可以帮助企业优化生产销售等经营活动,最后实现监管和智能商业流程优化的统一。

6.3 智慧城市大数据应用的政府服务领域

6.3.1 领域大数据分析

1) 政府领域大数据应用简介

(1) 国内外政府应用大数据现状。在政府服务领域,国内外一些先行者已经在运用大数据,通过多渠道的数据采集和快速综合的数据处理,增强治理社会的能力,实现政府公共服务的技术创新、管理创新和服务模式创新。大数据在政府服务领域的应用不仅使传统难题迎刃而解,更成为新时期应对新挑战、解决新问题的必然选择。

美国 2009 年面向市民提供数据开放网站,分为原始数据、地理数据、数据工具三类数据,在网站上有 40 万种原始数据文件,涵盖了农业、气象、金融、就业、交易、医疗、交通等;奥巴马认为,数据在未来将是领地权、领海权、领空权之外的另一种国家核心资产。美国将大数据提高到国家战略层面,形成全体动员的格局。

欧盟对大数据技术的应用也在计划之中,该技术可为欧盟 23 个最大的政府公共部门管理活动的成本提供 15%~20% 的下降空间,在未来 10 年,每年创造 1 500 亿~3 000 亿欧元的价值,并将公共部门的预计效率提高 0.5%。

伦敦通过大数据提高了服务的有效性,可以便捷地了解全市交通状况,并且把相关的数据和分析结果以可视化的方式上传到其门户网站,市民可以通过自身需要通过手机、车载等获取地图、文字、视频等信息。

西班牙首都马德里整合了警察、消防、医疗系统,使救援时间大幅度缩短,巡逻队、消防车、救护车能够在 8 min 内到达 81% 的突发事件现场。

新加坡通过大数据,提出了一站式的移动服务平台、一站式的政府数据平台;智能交通综合信息管理平台在预测交通流速和流量方面有高达 85% 的准确率,能通过有效地引导和干预,显著提升高峰时段的车辆通行效率。

我国各地政府部门大数据战略应用情况如下:

① 苏州通过覆盖城乡的信息化防控网络,在警力与人口配比不足万分之十的情况下,使打击处理案件数、刑拘转捕率、技术支撑率均为全市最高,实现了“以十抵万”的办案效率。

② 北京通过政务数据资源网开放数据,提供 400 多个涵盖旅游、教育、交通、医疗等数据包,可以下载或应用开发。

③ 上海出台大数据研究与发展的三年行动计划。

④ 广东率先出台大数据战略。

⑤ 顺德开展大数据时代的政务建设。

⑥ 山东成立农业大数据联盟。

⑦ 浙江通过大数据在高速公路上治堵。

(2) 政府服务领域大数据应用简介。大数据具有大量(Volume)、高速(Velocity)、多样(Variety)和价值(Value)的特征。大数据在政府服务领域的应用,主要体现在处理信息方式的变化:一方面,通过大数据包容性的应用,实现政府各部门之间、政府与市民之间的信息共享,提高政府各机构协同办公效率和为民办事效率,实现政府社会治理能力和公共服务能力的显著提升。另一方面,由于大数据改变了公众思考世界、处理问题的方式,人们更多地关注信息之间的关联度,以及关联之后产生的更多价值。因此,大数据通过处理和分析而被发掘出来的价值将成为公众服务的核心内容,政府服务模式的转变如图6-7所示。

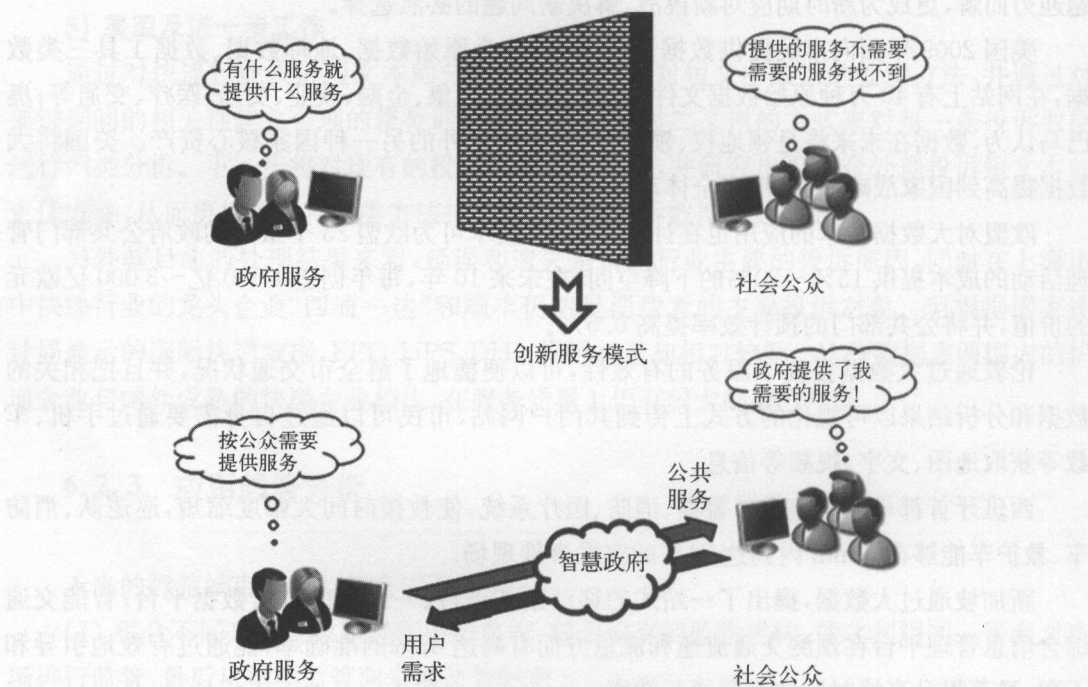


图 6-7 政府服务模式转变

如何利用大数据技术和理念,确保公众能够更多、更好地获取政府信息,为市民提供更为精准的信息推送服务,为政府决策提供更有力的支撑,成为大数据背景下建设服务型政府的重要内容。

2) 政府应用大数据特征分析

大数据能够在各个领域显著提高创新力、竞争力和产出率,相对于其他行业,政府服务领域应用大数据面临的困难最小、从大数据中获得的收益更多、价值潜力更大。这是因为,大数据的核心是数据、技术和思维,只有先占有巨量的数据,才能从中挖掘出巨大的价值。

我国政府应用大数据有如下特征：

(1) 专门的统计部门和工作队伍。国家统计局会定期开展人口普查和经济调查,大多数部委都设有发展规划司,很多单位都设有发展规划处,而财政、交通和气象等部门也掌握大量有关经济社会运行的数据。

(2) 民生和企业组织数据获取。政府工作关系着民生的方方面面,在日常行政过程中,也自然而然地积累了各类与社会生活息息相关的数据。政府还可以根据需求,要求企业、事业单位、行业协会提供各种数据。

(3) 政府应用大数据影响力大。政府不仅是大数据的受益者、大数据的占有者,更在建设大数据基础设施、培育大数据产业、培养大数据人才、完善相关标准和立法等方面负有至关重要的责任。尤其在我国的,政府在资源配置方面发挥着重要的作用,善于集中力量办大事,其强大影响力是带动大数据加速发展的优势所在。

(4) 政府管理需要模式转型。我国政府在大数据应用方面刚刚起步,要利用好大数据,不仅面临技术因素方面的困难,更需要一系列的大转型,如图 6-8 所示。把大数据的手段和方法引入管理领域,是实现管理智慧化的有效途径,也是大数据时代的必然要求。

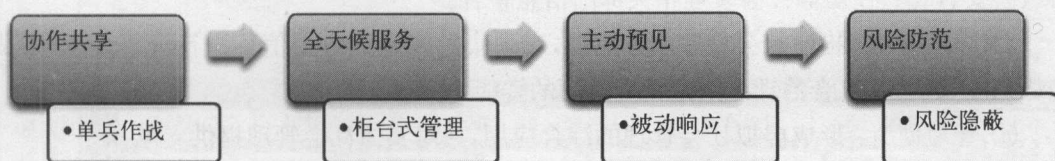


图 6-8 政府管理模式转变

创新社会管理和城市运营一体化,是智慧城市背景下政府管理与服务大数据新模式的典型代表,下面将对这两个领域展开详细介绍。

3) 创新社会管理应用

(1) 社会管理概况。

① 传统社会管理面临不少问题：

- 人员流动导致属地化管理面临困难
- 突发事件频发,应急处置压力大
- 部门分散导致社会管理中的信息共享和业务协同困难
- 网络繁荣让虚拟社会成为人们生活的重要部分,而对虚拟社会缺乏有效的管理

② 智慧城市建设需要创新社会管理：

- 智慧城市的核心理念是以人为本,创新社会管理也是体现以人为本的理念
- 智慧城市将促进信息共享和业务协同,这给构建一体化的社会管理体系提供了

基础

➤ 让生活美好,不仅在实体社会,也要体现在虚拟社会。营造良好的虚拟社会环境是创新社会管理的重要途径之一

③ 创新社会管理需要大数据技术手段支撑:

- 需要技术手段更及时地采集社会动态信息
- 需要技术手段对基于网络的虚拟社会进行监管
- 需要技术手段整合分散在各部门、各系统的社会管理信息,实现信息共享和业务协同
- 需要技术手段对海量的社会动态信息进行综合分析、研判,以形成有效的预警机制
- 需要建立一些数据模型,提升社会管理的智能化程度

(2) 创新思路和概念模型。应用大数据和云计算等新技术,为社会管理提供一体化平台和整合的应用服务,并通过区级应用示范,探索实体社会管理与虚拟社会管理相结合的社会管理一体化模式。

① 动态感知:对现实社会和虚拟社会中的人员、场所、设备设施、活动、舆论等信息进行动态采集。

② 互联互通:实现区、街道、镇、居委、商务楼宇等社会综合管理部门的信息共享,并与市级公安、司法、住房等社会管理相关部门信息整合。

③ 应用智能:构建社会管理业务模型,对采集的海量大数据信息进行分析处理,形成趋势判断,为社会综合治理提供预警和处理的智能化手段。

④ 管理创新:形成虚拟社会管理的综合应用,并对实体社会管理提供支撑。

创新社会管理一体化平台将全面整合全区突发事件、问题隐患和常态化业务信息,实现多层次信息共享和业务联动,如图 6-9 所示。

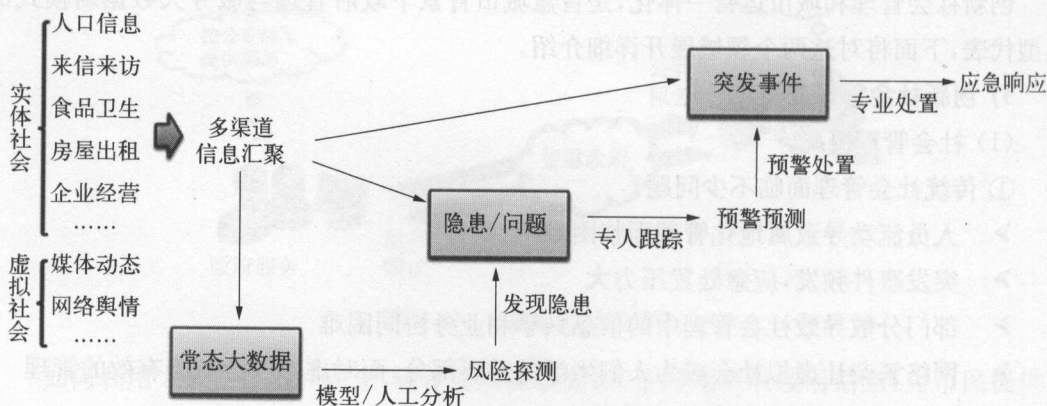


图 6-9 创新社会管理一体化平台概念模型

(3) 创新社会管理总体框架(图 6-10)。基于一体化平台的创新社会管理基本框架包括:

- 全面联动的社会管理机构
- 以人为本的社会建设和保障体系

- 完整有效的社会管理的服务和协调机制
- 一体化整合的社会管理信息化支撑平台
- 多方位协同的社会管理信息化综合应用



图 6-10 创新社会管理总体框架

(4) 创新社会管理建设内容。

① 社会管理一体化平台(PaaS 模式)。

- 综合数据管理：实现区、街道、镇、居委、商务楼宇等社会综合管理部门的信息统一管理，并与市级相关的公安、司法、综治、房管等多个部门信息共享，建立数据关联关系
- 信息共享服务：提供跨系统信息调阅、视频共享、信息分发等服务
- 应用集成服务：提供统一安全认证、流程协同引擎等应用支撑服务
- 运行管理服务：对各接入系统运行情况进行统一监管，提升数据质量和整体系统可用性

② 社会管理一体化应用服务(SaaS 模式)。

- 社会管理信息动态汇聚：通过智能终端、数字化监控等物联网设备，动态采集特定人员、重点场所等动态信息；通过整合单位热线、委办局应用系统实现条线信息的汇聚；通过网络数据监测等方式，动态采集舆情、党建等虚拟社会活动信息

► 城市常态化协同联动: 提供行政执法、隐患排查、业务监督、综合服务和舆论导向的管理功能, 加强常态化行政管理与执法联动工作的开展

► 城市应急响应处置: 提供面向突发公共事件的快速响应, 形成全流程、智能化、协同化的处置支撑系统, 满足快速的事件处置和面向上级的信息报送

► 多维度领导辅助决策: 通过统计分析、调度研判、预警预测、效能评估等系统的建设, 满足领导对城市运行情况整体把握的需求, 提供辅助决策支撑

► 多渠道信息综合服务: 提供面向不同渠道和用户的信息服务手段, 提供指挥中心、管理服务中心和网络、短信、传真等渠道服务

③ 社会管理关联系统集成与总集成。与公安、司法、工商、房地产、网格化管理中心等社会管理部门进行系统集成, 实现包括视频共享、数据整合、跨部门信息调阅、协同处置等功能。

④ 社会管理一体化系统配套体系研究。

► 社会管理一体化运营模式设计

► 社会管理一体化系统标准规范

4) 城市运营中心应用

2012年, IBM为巴西里约热内卢打造了智慧城市运营中心, 将30多个城市管理部门数据统一于城市运营管理系统。在运营中心的大屏上, 播放里约热内卢城市动态监控视频, 包括各个地铁站、主要路口的交通状况, 交通事故处理情况, 天气预测系统预报的城市未来几天的降雨情况, 停电处理情况, 以及其他城市问题处理及其进展等状态。我国智慧城市应用同样需要一体化的城市运营中心。

(1) 城市运行管理存在的问题。城市运行涉及城市管理的方方面面。从行业上分, 包括交通、市政、公共安全、水利、环境保护、医疗防疫、电力、气象、食品安全、应急指挥等众多专业部门; 从层级上分, 分为市、区县、街道、社区等不同管理层的指挥机构。对以上各类运行工作进行监控, 及时展现它们的运行状态、问题处理情况以及其他方面的情况, 是城市管理者进行指挥的前提。但目前国内各城市在运行管理过程中, 都面临一些类似的问题。

① 缺少信息的整合和共享应用。城市相关部门依据各自的管理职能建设了各自的业务信息管理系统, 但由于缺少综合协调、各自利益为重等原因, 全市城市运行相关信息资源未能得到有效整合, 系统之间信息也不能共享, 或者只是有限共享, 已建的系统成为信息孤岛。

② 行业信息化发展不均衡。城市运行指挥人员面临的主要问题是得不到准确及时的信息支持。由于各种复杂原因, 城市运行相关专业部门信息化建设水平参差不齐。部分机构信息化水平较高, 能够及时提供反映本专业运行状态的信息, 另一些部门则还基本处于手工统计阶段, 运行状态信息往往要延迟很久才能得到。而且, 由于长久以来缺乏一种及时汇总和形成城市运行状态信息的机制, 各专业部门统计的信息往往没有被充分利用。最终的结果是, 领导需要的时候, 得不到及时有效的信息支持, 而一些有价值的信息, 又被束

之高阁,得不到充分利用。

③ 信息的深层次挖掘利用还比较欠缺。当前部门所建信息系统基本只用于为本单位提供日常业务支撑服务,而未对信息的利用进行深层次开发,无法为城市整体运行管理智慧化提供数据支撑,使系统的效能大打折扣。

对于城市管理者而言,缺乏反映整个城市宏观运行情况的适当信息支持。城市运行涉及诸多领域,这些专业的统计数据,倘若不经过专家的解释,非专业人士往往很难搞清楚这些数据的意义和其所蕴涵的信息。事实上,决策人员并不是特别关注数据本身,而其中蕴涵的信息才是他们特别需要掌握的东西。

因此,需要从各专业的统计数据中,抽取出有价值的信息,以直观的方式展现给决策者,并综合各专业的运行情况,形成全市的城市运行情况,从而让决策者能够一目了然地了解整个城市的运行态势和任意一个局部专业的运行态势。

④ 信息传递不畅,综合协调机制不足。城市管理者得不到及时准确、形象直观的信息支持,除了信息本身的不足外,信息沟通传递不畅也是一个重要原因,其结果是导致了对城市运行指挥协调的难度加大,降低了领导决策的科学性和准确性。尤其像交通和市政保障运行这样的工作,它们是城市运行各项工作任务的重中之重,一旦出现情况,不仅无法迅速得到处理,还特别容易造成全局性影响。由于这些领域存在分管部门多、协调头绪多、相互干扰大、隐患排查难等现象,因此,迫切需要构建城市日常运行指挥协调机制及配套的信息沟通机制,统一协调城市日常运行指挥中面临的问题。

(2) 智能城市运营中心总体思路。智能运营中心的建设以“智慧的城市运行管理”为目标,在现有信息化建设基础上,通过资源整合将全市各部门业务系统以及行政服务部门部署在不同设备、不同系统采集到的城市基础信息融合成统一的数据视图,在此基础上进行智能分析和仿真预测,通过数据的智能分析得出城市运行相关的各类智慧信息,充分挖掘城市运行中的各类信息的价值,为城市管理者提供决策支持,把城市管理模式从被动应对事件转变成有预见性地主动、提前规避问题,使城市能够更好地预见问题、应对危机和管理资源。

(3) 智能运营系统总体内容。智能运营中心将作为智慧城市各个运行环节和各类运行资源的整合中心及物理门户,全面实现智慧城市运行的监控、协调、指挥、展示,其大脑和枢纽是运行管理与服务平台。

该平台的实现依托于全市各部门以及政务数据中心数据资源,在其基础上进行整合、分析、挖掘,并以数据的整合逐步推动管理的整合,形成一个智慧城市运营管理的核心平台。

该平台的核心价值是突破城市传统“按部门”的管理维度,提供城市管理者从构成城市的要素的全新管理视角,如整合多部门数据,提供以事件、人、道路、车辆、自然资源、公共设施、企业等为中心的 360°全方位视图;通过 GIS 空间地理信息系统,将全市政治、经济、文化等各要素,水、电、交通等其他各类海量的服务资源信息,结合地理信息,准确、实时、高效、

生动、直观地进行展示,为领导决策和民众服务;通过各专业部门间城市数据的共建共享,构建各部门间的协作应用;对海量数据进行分析利用,构建城市运行和状态数据的监测,在数据汇聚基础上进行综合分析和预测,为城市运行提供智慧管理的决策依据。具体应用建设内容包括综合指标分析展示功能、城市资源综合展示及管理功能,涉及交通城市管理、环保等领域智能协同功能、数据传输与交换子平台、智能门户、基于 DLP 的视图功能、外网智慧政务(综合政务平台)建设等。

6.3.2 典型应用场景

1) 社会管理信息动态汇聚(图 6-11)

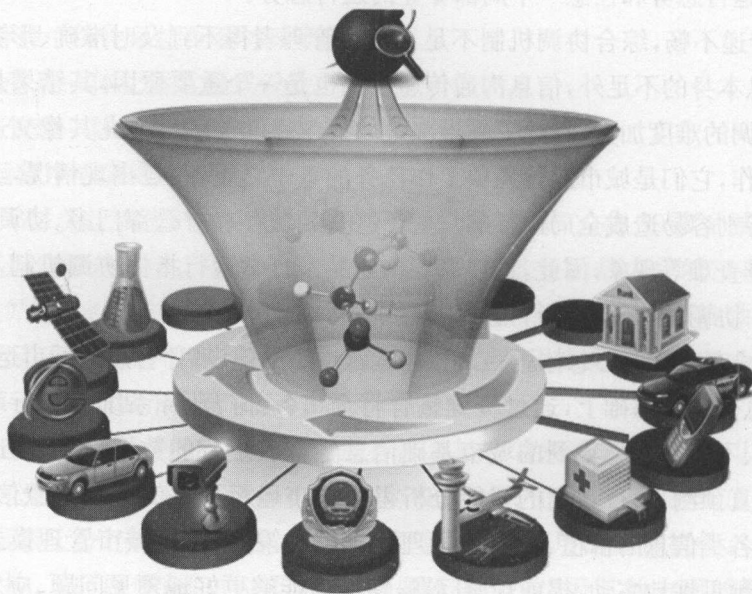


图 6-11 社会管理信息动态汇聚场景

(1) 场景目标。搭建全面覆盖城市运行、社会管理和虚拟社会活动的信息综合采集系统。

(2) 场景内容。

① 城市运行信息采集:通过智能终端、数字化监控设备联网,动态采集特定人员、重点场所等动态信息。

② 社会管理信息采集:通过整合单位热线、委办局应用系统实现条线信息的汇聚。

③ 虚拟活动信息采集:通过网络数据监测等方式,动态采集舆情、党建等虚拟社会活动信息。

(3) 应用效果。

① 实现对城市运行和社会管理信息的信息管理汇聚。

② 实现对虚拟活动信息的集中分析,进而实现对城市运行和社会管理影响的衍生分析。

③ 形成对城市整体运行的通盘信息把握。

2) 城市常态化协同联动

(1) 场景目标。提供行政执法、隐患排查、业务监督、综合服务和舆论导向的管理功能,加强常态化行政管理与执法联动工作的开展,如图 6-12 所示。

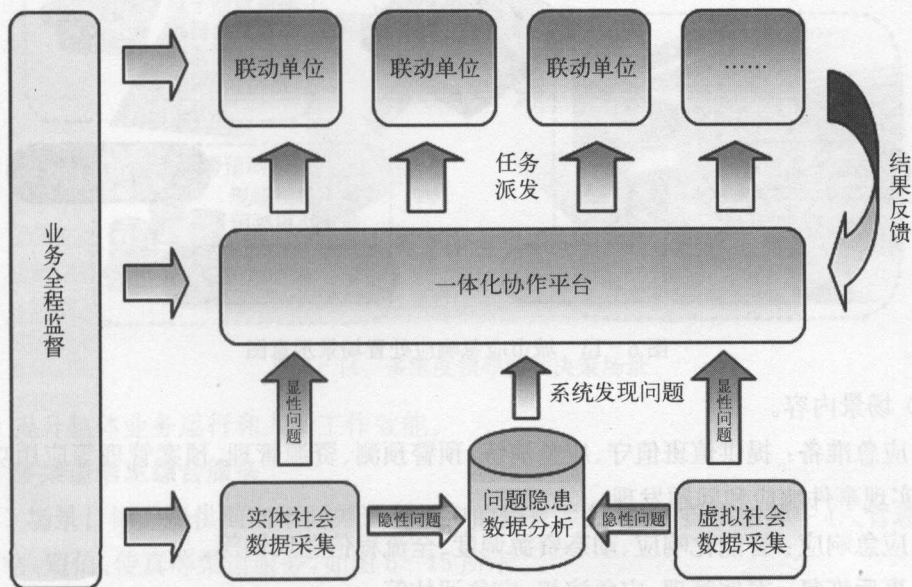


图 6-12 城市常态化协同联动场景

(2) 场景内容。

- ① 行政执法：面向公安、安监、卫监、文化、教育、工商的城市管理执法业务支持。
- ② 隐患排查：对安全生产和社会不稳定因素进行排查,提供风险预警。
- ③ 业务监督：对各项业务的处置过程实现监管,支持督查督办和业务监管。
- ④ 综合服务：提供面向不同成员单位的信息服务和办事流程服务。
- ⑤ 舆论导向：对重大事件和虚拟社会事项(党建、舆情)提供信息服务和舆论导向支撑,提升精神文明建设的引导力度。

(3) 应用效果。

- ① 常态管理联动化。
- ② 风险隐患及时显性化。
- ③ 业务过程可监督追溯。
- ④ 意识形态管理水平提升。

3) 城市应急响应处置

(1) 场景目标。提供面向突发公共事件的快速响应,形成全流程、智能化、协同化的处

置支撑系统,实现快速的事件处置和面向上级的信息报送,如图 6-13 所示。

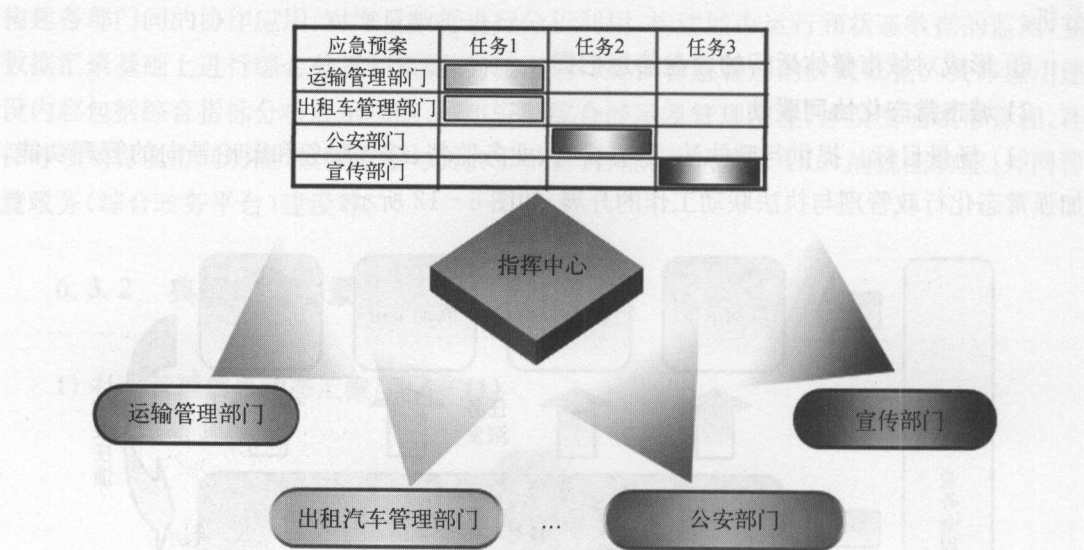


图 6-13 城市应急响应处置场景示意图

(2) 场景内容。

① 应急准备: 提供值班值守、应急演练、预警预测、资源管理、预案管理等应用功能,能够快速实现事件响应和问题发现。

② 应急响应: 智能化响应、动态资源调度、全流程信息反馈等。

③ 事后恢复: 案例管理、应急演练、应急评估等。

(3) 应用效果。

① 全流程应急管理和响应。

② 一体化资源调度和管理。

4) 多维度领导辅助决策

(1) 场景目标。通过统计分析、调度研判、预警预测、效能评估等系统的建设,满足领导对城市运行情况整体把握的需求,提供辅助决策支撑,如图 6-14 所示。

(2) 场景内容。

① 统计分析: 提供多维度统计分析工具,以图表等形势支撑对业务情况、发展趋势的研判需求。

② 调度研判: 领导可通过系统进行指令发布、信息分析、讨论会商。

③ 预警预测: 系统提供信息预警和多数数据统筹预测,提升数据利用水平。

④ 效能评估: 引入行政效能模型,能够对整体工作效能进行量化评估。

(3) 应用效果。

① 领导通过系统快速清晰通览全局。

② 系统提供智能化的研判、预警、趋势研判辅助功能。



图 6-14 多维度领导辅助决策场景

③ 提升整体业务运行和人员工作效能。

5) 多渠道信息综合服务

(1) 场景目标。提供面向不同渠道和用户的信息服务手段,提供指挥中心、管理服务中心和网络、短信、传真等渠道服务,如图 6-15 所示。



图 6-15 多渠道信息综合服务场景

(2) 场景内容。

① 指挥中心：提供统一指挥中心，通过视频监控、接处警、一体化社会管理应用建设实现常态化管理和应急业务支撑，提升城市管理水平。

② 社会管理服务中心：整合已有的市民服务中心、社区服务中心的信息服务功能，实现社会管理信息的综合发布。

③ 信息服务渠道：市民可通过手机、电视媒体、短信等多种渠道获取各类公开信息，在不同单位间能够通过无纸化方式完成信息高效流转。

(3) 应用效果。

① 指挥调度集中化。

② 管理服务多渠道化。

③ 信息权威性、准确性保障。

6) 城市运行监测指挥业务场景

(1) 场景目标。智能运营中心是在城市管理平台实现日常城市管理问题的早发现、早报告、早投入、早部署、早处置这一“事后处置”的基础上对“事前防范”进行的积极探索和尝试，如图 6-16 所示。

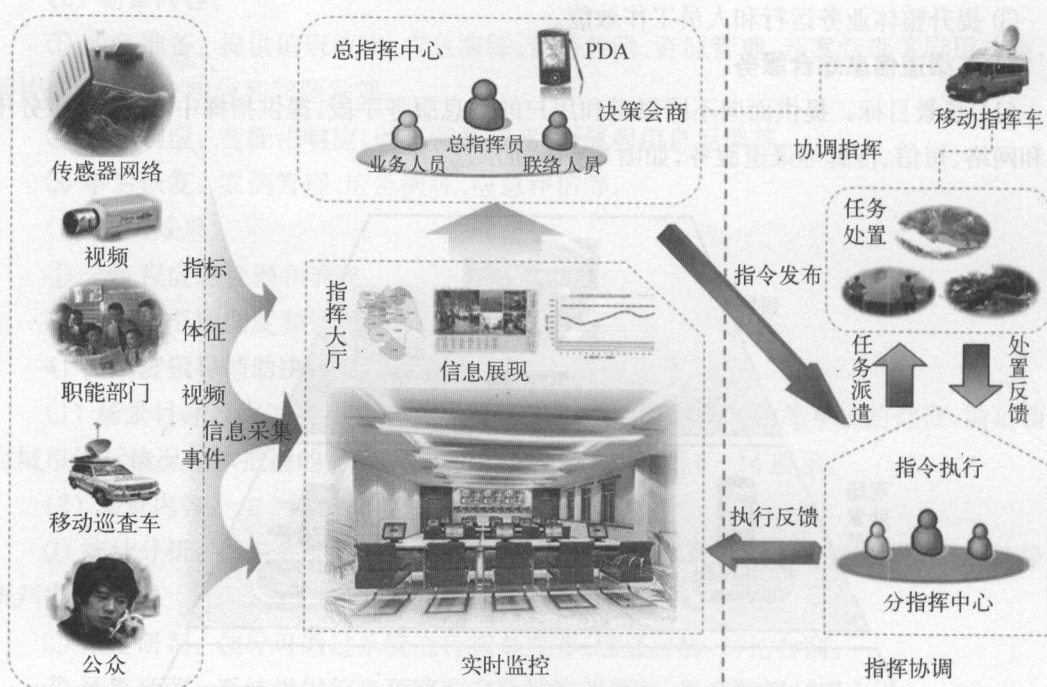


图 6-16 城市运行监测指挥业务场景

(2) 场景内容。系统通过信息填报、视频监控、传感器监测、呼叫服务等不同手段，获取城市运行关键指标、重大事件等城市运行信息，在指挥大厅进行实时展现，及时对城市运行态势进行分析预测，为管理部门和领导提供决策信息支撑，做到事前防范和干预及事后的

指挥和处置,实现全方位的城市运行管理。

(3) 应用效果。通过多维多色态势模型进行直观可视化展现,帮助领导及时掌握城市运行体征,准确“把脉”城市运行态势,实现科学决策,从而在宏观上尽早预警城市管理问题的发生,实现将城市管理关口前置,防患于未然。同时,实现城市重大突发事件处置过程中的领导辅助决策和指挥调度,提高政府领导的城市管理问题决策和指挥调度能力,提高现代化城市管理水平。

6.3.3 应用趋势分析

1) 政府应用大数据总体趋势分析

我国政府在大数据方面的应用刚刚起步,除了技术上的突破,政府服务领域应用大数据的总体趋势在于管理的突破与创新,并与智慧城市建设相结合,提升政府服务和城市管理水平。

(1) 政府管理与服务应用领域大数据应用趋势。

① 实现信息透明与共享,使外部利益相关者(如公民与企业)和内部利益相关者(如政府雇员和政府机构)都能提高自身的工作效率,产生积极的经济社会综合效益。

② 通过评估政府部门的绩效,增强内部竞争,激励工作表现,提高公共建设效率,提升行政服务质量,降低政府的管理成本。

③ 通过人口细分和定制政策,增强公共服务的针对性,提高工作效率和公众满意度,减少开支。改变传统政府部门为所有公民提供相同服务的现状,满足公众多元化、个性化的服务需求。

④ 由数据分析和政务智能替代或辅助人工决策,在纷繁复杂的数据中自动识别出不一致、错误和虚假的信息,减少出错成本和福利管理中的诈骗,缩小税收缺口。

⑤ 引导公共部门内部和外部的创新,例如,商业、非营利机构、第三方通过开发出大数据工具和分析,对公共服务进行反馈,为改善现有的方案提出建议,从而为政府部门创造新的价值。

(2) 大数据在城市管理中的应用趋势(智慧城市)。智慧城市可实现通过协调不同部门的资源,以提供更快、更有效的响应。大数据的智慧化帮助管理者跨部门决策、协调,提高公共服务交付效率,减少应对突发灾害的响应时间。

大数据是智慧城市实现智慧化的关键支撑。从政府决策与服务到居民衣食住行,从城市的产业布局规划到城市的运营和管理,大数据遍布城市的方方面面,成为驱动城市管理发展的引擎。与此同时,智慧城市物联化、互联化和智能化的特点本身也会催生海量数据。

智慧城市是面向城市整体的综合运营系统,信息覆盖更加全面,信息感知更加深入,信息共享和互联更加通畅,对海量数据的处理机制更加完善,并且能够分析预测、辅助决策。

在不远的未来,数据将成为维护城市运行的基本要素,政府管理部门建立立体的大数据生态系统,让城市更加宜居,让居民生活更加幸福。

(3) 政府应用大数据与国家规划相结合。在我国政府最新出台的《城镇化规划(2014—2020)》中,对城市管理和公共服务作出了规定:统筹城市发展的物质资源、信息资源和智力资源利用,推动物联网、云计算、大数据等新一代信息技术创新应用,实现与城市经济社会发展深度融合。强化信息网络、数据中心等信息基础设施建设。促进跨部门、跨行业、跨地区的政务信息共享和业务协同,强化信息资源社会化开发利用,推广智慧化信息应用和新型信息服务,促进城市规划管理信息化、基础设施智能化、公共服务便捷化、产业发展现代化、社会治理精细化。增强城市要害信息系统和关键信息资源的安全保障能力。

大数据在我国政府管理与服务中的应用,也需要与以上规划思路相结合。

2) 创新社会管理应用大数据趋势分析

- (1) 在传统社会管理的基础上,引入虚拟社会管理,并实现两方面密切关联。
- (2) 一体化平台促进社会管理各部门和环节的共享与协同。
- (3) 将大数据、云计算和物联网等新的技术和服务模式与创新社会管理的要求相结合。
- (4) 建立社会管理的业务模型,提高数据综合分析和预警能力,提升社会管理应用的智能化程度。

3) 城市运营管理应用大数据趋势分析

- (1) 全面汇聚、有效利用城市运行的各级各类信息。智能运营中心应实现城市运行信息的全面整合与共享;管理人员可以通过来自城市多个部门、多个单位甚或各类传感器的各类信息,以最有效的方式了解城市运行的方方面面。
- (2) 及时掌握、智能预测城市运行可能出现的问题。智能运营中心能够在海量信息积累的基础上,实现对城市运行中不良情况的智能判断和预测;管理人员可以通过分析城市运行的状态是否正常,预测可能发生的问题和事件,提前预防,主动应对,将事件对城市运行产生的负面影响降至最低。
- (3) 快速响应、协调资源推进城市智慧管理和服务的一体化。智能运营中心能够为市领导和相关部门提供高价值的城市运营信息,为城市管理的协同作业提供一个统一的门户和平台;城市运行的各管理和参与部门在市领导的统一领导下,实现一体化作业,通过高效率的资源协调和协同作业,以一个整体来解决城市运行中的问题和处理应急事件;为市民提供相应的服务。

6.4 智慧城市大数据应用的基础设施领域

6.4.1 领域大数据分析

城市基础设施(Urban Infrastructure)是城市生存和发展所必须具备的工程性基础设

施和社会性基础设施的总称,是城市中为顺利进行各种经济活动和其他社会活动而建设的各类设施的总称。它对生产单位尤为重要,是其达到经济效益、环境效益和社会效益的必要条件之一。工程性基础设施一般指能源系统、给排水系统、交通系统、通信系统、环境系统、防灾系统等工程设施。社会性基础设施则指行政管理、文化教育、医疗卫生、商业服务、金融保险、社会福利等设施。我国一般讲城市基础设施多指工程性基础设施。

大数据既是社会热点,也是一项非常复杂的数据工程。必须要加强顶层设计,开展深度研究。既要研究基础算法和应用方向,也要结合系统建设实际研究实施方案和实施策略。

6.4.2 典型应用场景

典型应用如国家电网在全系统实施的 SG186 工程,自工程建设实施以来,营销相关信息系统已经覆盖了 3 亿多电力用户,各省(市)公司已经累计超过 1 000 TB 的数据。数据范围已经从传统的营销业务应用系统扩展到用电信息采集系统、电能服务管理平台、电动汽车充换电运营管理系统、计量生产调度平台、分布式电源运营管理系统、95598 客服系统、实时费控系统、营销 GIS 系统等相关业务系统。营销自动化和信息化工作的推进,为营销大数据工作的开展提供了必要条件^[17]。但海量的营销数据不仅没能成为挖掘知识、提炼数据价值的宝藏,反而成为数据存储和容灾恢复的负担,故急需开展营销大数据的研究工作。开展营销大数据应用也是提升客户服务水平、支撑决策智能化、管理精细化的要求。

首先,开展营销大数据应用是提升客户服务水平的需要。随着社会经济发展,用电客户的大量增长,使得营销的客户服务工作面临新挑战。多元化的客户需求必然要求客户服务渠道的多元化、服务内容的多元化及服务水平的差异化和个性化。只有通过大数据应用,分析用户的基础档案、用电行为、缴费轨迹等,才能对客户合理分群、科学分类。再通过大数据的分析,制定合理的服务套餐,提供差异化、个性化的用电服务。

其次,开展营销大数据应用是实现管理精益化的需要。随着 SG186 工程实施及大营销体系建设纵向深入开展,管理精益化已经成为新型电力营销模式的新特征。在此过程中,业扩流程简化、电费回收、计量资产管理均取得了较高的管理成效。依照传统的管理手段和管理方式已经难以提升管理水平。只有通过大数据分析手段,开展业扩流程分时段时长分析,研究该时长的用户分布、用户行为分布、关键环节分布,继而挖掘管理短板,制定相应的管理策略,改变管理手段,切实提升管理效益。

再次,开展营销大数据应用是实现业务决策智能化的需要。随着营销自动化建设的不断深入。从传统的营销业务应用、用电信息采集等业务管理系统到支撑新型业务发展的电动汽车充换电系统、电能服务管理平台和分布式电源运营管理系统,再到整合数据资源的基础数据平台和总部业务管理平台,营销自动化水平已经达到一个新的台阶。传统的经验主义的市场分析预测、有序用电规划、电能表计需求上报、营业厅选址规划等业务决策已不

合时宜。只有通过大数据研究分析,研究用户运行轨迹,聚焦用电客户的活动范围,才能开展营业厅的选址规划;才能建立相关的市场分析模型、有序用电模型、电能表需求量计算模型以支撑营销决策的智能化。

6.4.3 应用趋势分析

与业务系统建设相比,城市基础设计的大数据应用见效周期长,往往需根据业务对构建的数据模型进行多轮的迭代修正后才具备实用化条件,并需在后续应用中继续进行迭代修正完善,因此不能像建业务系统一样要求立竿见影的成效。

◇参◇考◇文◇献◇

- [1] 周雪忠,吴朝晖. 文本知识发现: 基于信息抽取的文本挖掘[J]. 计算机科学,2003(1): 63-66.
- [2] 王细薇,沈云琴. 中文短文本分类方法研究[J]. 现代计算机,2010(7): 28-31.
- [3] 孟宪军. 互联网文本聚类与检索技术研究[D]. 哈尔滨: 哈尔滨工业大学,2009.
- [4] 邓莎莎. 支持决策研讨的文本分析方法研究[D]. 上海: 上海交通大学,2013.
- [5] 成希瑶. 我国快递行业发展现状及对策研究[J]. 中国市场,2011,32: 17-18.
- [6] 绍络姆·韦斯,霓廷·因杜里亚,张潼. 预测性文本挖掘基础[M]. 赵仲孟,侯迪,译. 西安: 西安交通大学出版社,2012.
- [7] Mehran Sahami. Generating neural networks through the induction of threshold logic unit tree [J]. Engineering Applications of Artificial Intelligence, 1996,9(2): 129-136.
- [8] Christos P, Prabhakar R, Hisao T, et al. Latent Semantic Indexing: A probabilistic analysis (Postscript) [J]. Proceedings of ACM PODS, 1998.
- [9] Thomas H. Probabilistic Latent Semantic Indexing [C]. Proceedings of the Twenty-Second Annual International SIGIR Conference on Research and Development in Information Retrieval, 1999.
- [10] 李振星,徐泽平,唐卫清,等. 全二分最大匹配快速分词算法[J]. 计算机工程与应用,2002(11): 106-109.
- [11] 湛燕,陈昊,袁方,等. 基于中文文本分类的分词方法研究[J]. 计算机工程与应用,2003,23: 87-88,91.
- [12] 费洪晓,康松林,朱小娟,等. 基于词频统计的中文分词的研究[J]. 计算机工程与应用,2005(7): 67-68,100.
- [13] 彭泽映,俞晓明,许洪波,等. 大规模短文本的不完全聚类[J]. 中文信息学报,2011(1): 54-59.

- [14] 张志飞,苗夺谦,高灿. 基于 LDA 主题模型的短文本分类方法[J]. 计算机应用,2013(6): 1587 - 1590.
- [15] 孙建旺,吕学强,张雷瀚. 基于语义与最大匹配度的短文本分类研究[J]. 计算机工程与设计,2013(10): 3613 - 3618.
- [16] 蔡月红,朱倩,孙萍,等. 基于属性选择的半监督短文本分类算法[J]. 计算机应用,2010(4): 1015 - 1018.
- [17] 程静. 智能电网建设下 SG186 电力营销业务系统的管理应用与思考[J]. 中小企业管理与科技,2011(10).

第7章

大数据在智慧城市中的应用相关问题分析

本章主要内容包括产业分析、业务模式、指标体系和安全隐私四个部分。其中,产业分析主要介绍大数据应用对智慧城市发展的意义、大数据在智慧城市的国内外应用现状和市场规模与发展趋势;业务模式主要介绍大数据业务的商业模式的涵义、大数据业务的商业模式类型和大数据业务运营中的问题及对策;指标体系主要介绍指标体系概述、智慧城市的评价原则和评价指标;大数据的安全隐私主要介绍物联网的隐私维护与信息安全以及云计算的隐私维护与信息安全。

7.1 产业分析

7.1.1 大数据应用对产业发展的意义

智慧城市的发展理念为我国城市化带来了新的契机。智慧城市通过把互联网、云计算、大数据等新一代信息技术充分运用在城市的各行各业,实现全面透彻的感知、宽带泛在的互联、智能融合的应用,促进生产方式、消费模式变革和创新,从而带动产业的发展。由城市数字化到城市智慧化,关键是要实现对数字信息的智慧处理,其核心是引入大数据处理技术。大数据的应用在促进城市可持续发展的同时,推动产业转型升级;对产业发展具有重要意义。

1) 推动产业转型升级,促进信息消费,推动城市可持续发展

当前,我国很多城市的支柱产业都是资源密集和劳动密集型产业,缺乏核心技术和优势品牌。随着资源逐渐枯竭,资源依赖型城市发展无法持续,转型迫在眉睫。城市老龄化加快,人口红利减少,劳动力成本不断增加,劳动密集型产业逐渐向内陆或是劳动力更廉价的地区转移,急需发展新型产业支撑经济发展。如何促使现有产业转型升级,培育一批具有核心竞争力的产业,促进经济健康可持续发展是摆在各级政府面前的一大难题。大数据的合理化应用可为智慧城市建设、促进城市产业转型升级和经济健康发展提供关键支撑。

一方面,大数据在智慧城市建设中将推动政府、企业数据资产的开放和共享,从而进一步促进信息消费。2013年7月12日,国务院总理李克强主持召开国务院常务会议,研究部署加快发展节能环保产业、促进信息消费、拉动国内有效需求、推动经济转型升级。信息消费与大数据技术的发展与应用密切相关,是极具发展潜力的消费领域。由于信息技术的快速发展,智能手机、平板电脑等电子产品的广泛应用,以及云计算在多个行业中的广泛布局,使数据信息逐渐演变成为一种促进经济发展的新能源。大数据技术作为对信息进行采

集、分析、挖掘、预测的工具,成为扩大信息消费规模的重要引擎,数据技术不仅为互联网行业带来翻天覆地的变化,未来更会在政府决策、智慧城市、互联网金融、外贸服务等方面扮演重要角色,而这些领域的背后,均存在着不同模式信息消费的身影。

另一方面,大数据的应用可促使丰富数据资源和技术优势的IT企业向互联网信息服务企业转型。大数据渗透于人们生活的方方面面中,并在一定程度上关联着中国现代化发展进程。以大数据为主要驱动力的企业应当将目光放长远,积极将大数据技术拓展至更多行业中去,创造出更多更好的产品与服务,在推进更多传统行业转型的同时,也为企业自身开拓出一片商业新蓝海。

第三方面,通过促进物联网、云计算等战略性新兴产业的应用和发展,带动信息产业在国际竞争中的超越。大数据的应用将是推动工业化和信息化深度融合的重要力量,企业高速发展的引擎。大数据协助企业对内部数据资产进行深度分析和挖掘,准确定位客户需求,更有针对性地开发改善产品,提高创新能力和服务质量,提升客户和受众体验,在竞争中获取优势。大数据为企业制定战略提供帮助,为快速精确决策、降低运营成本、提高业务效率和拓展新型商业模式提供有效支持,从而帮助企业超越竞争对手,抢占市场先机,提高市场份额。大数据在帮助企业不断创新发展的同时也提升了各类产业的总体水平,带动城市经济快速发展。

2) 提高政府决策水平,提升科学治理能力

智慧城市的建设带来数据量的爆发式增长,海量数据就像血液一样遍布智慧交通、智慧医疗、智慧生活等城市领域的方方面面,传统依靠经验治理城市的模式无法满足智慧城市建设和运营的需要。大数据使政府实现科学治理成为可能。大数据在智慧城市中的应用可以为政府在城市各个领域提供强大决策,优化行政资源,降低管理成本,提高管理效率,支持推动城市科学治理,使公共支出产生最大的效益。

大数据将进一步提高决策的效率,提高政府决策的科学性和精准性,提高政府预测预警能力以及应急响应能力,节约决策的成本。建设大数据中心,加强政务数据的获取、组织、分析、决策,通过云计算技术实现大数据对政务信息资源的统一管理,依据法律法规和各部门的需求进行政务资源的开发和利用,可以提高设备资源利用率、避免重复建设、降低维护成本。

借助大数据,还能逐步实现立体化、多层次、全方位的公共服务体系,推进信息公开,促进网上办事实时受理、部门协同办理、反馈网上统一查询等服务功能,加快推进智慧城市建设和移动服务新模式的初步应用,不断拓展个性化服务,进一步增强政府与社会、老百姓直接的双向互动、同步交流。以财政部门为例,基于云计算、大数据技术,财政部门可以按需掌握各个部门的数据,并对数据进行分析,作出的决策可以更准确、更高效。另外,也可以依据数据推动财政创新,使财政工作更有效率、更加开放、更加透明。在市场监管方面利用大数据整合信息,将工商、国税、地税、质监等部门所收集的企业基础信息进行共享和比对,通过分析,可以发现监管漏洞,提高执法水平,达到促进财税增收、提高市场监管水平的

目的。

在城市规划方面,基于城市网格化的管理需要一个统一协调的管理信息整合,各类基础资源和信息都应该是共享的,大数据可以实现这一点。通过充分利用大数据的各类资源,发挥城市网格化管理效用,达到最大程度的共享应用,以提升城市和社区的服务质量、提高服务能力、加强服务管理,创建服务型社会,使城市管理工作和社区服务水平迈上更高的台阶。通过对城市地理、气象等自然信息和经济、社会、文化、人口等人文社会数据信息的挖掘,可以为城市规划提供强大的决策支持,强化城市管理服务的科学性和前瞻性。围绕环境、交通、医疗、教育等,构建城市日常运营以及应急联动指挥响应管理平台;在安防与防灾领域,通过大数据的挖掘,可以及时发现人为或自然灾害、恐怖事件,提高应急处理能力和安全防范能力。在舆情监控方面,通过网络关键词搜索及语义智能分析,能提高舆情分析的及时性、全面性,全面掌握社情民意,绘制社会不同时段情绪波动的实时色彩图,监控社会情绪,提高公共服务能力,应对网络突发的公共事件,打击违法犯罪,构建主动式虚拟社会管理体系和管理模式;在环保领域,通过实时监测水中各项参数,结合卫星遥感等多源异构数据,分析环境生态变化趋势;在交通管理方面,通过对道路交通信息的实时挖掘,能有效缓解交通拥堵,并快速响应突发状况,为城市交通的良性运转提供科学的决策依据。

目前我国基于大数据的信息共享建设取得了一定的成效,但是,从总体来看,跨部门的信息资源利用系统仍局限在小部分的政府业务范围,而且应用的深度和广度还远远不够,不能满足当今社会发展的要求。因此,我们需要顺应大数据这个趋势,通过推动大数据在智慧城市的应用,提高整个社会对信息资源的开发利用,从而提升政府的政府决策水平,提升科学治理能力。

3) 完善公共服务,优化服务质量

随着城镇化步伐的加快,城市人口迅速增长。城市医疗卫生、教育等资源严重不足且分布不均。以市民需求为导向,在教育、卫生、社区生活服务等公众关注度高的民生领域,提供覆盖、易使用的优质社会事业与公共服务,是智慧城市发展的重要方面。大数据在公共服务中的应用可以优化公共服务机构资源配置,提高服务质量,提升服务满意度。一方面,政府利用大数据技术对积累的海量历史数据进行挖掘利用,可以提供更有深度的公共服务。另一方面,政府可以通过对卫生、环保等领域的大数据实时分析,提高危机的预判能力,为实现更好、更科学的危机响应提供技术基础。如在医疗卫生领域应用,通过全面分析病人特征数据和疗效数据,比较多种干预措施的有效性,找出针对特定病人的最佳治疗途径;通过对匹配病源情况的数据筛查,构建疾病预防、医疗数据系统,准确找到感染群体、规模特征。提高医疗过程数据的透明度,可以使医疗从业者、医疗机构的绩效更透明,间接促进医疗服务质量的提高;根据医疗服务提供方设置的操作和绩效数据集,计算成本、质量和绩效,促进医疗服务机构提供更好的服务,从而更有竞争力。在交通系统,随着汽车工业的发展,车辆保有量的不断攀升,车与路、车与环境之间的矛盾日趋加剧,诸如交通堵塞、事故增多、能源浪费和环境污染等问题的恶化,需要通过对历史以及现在的车辆情况、路网情况

的实时大数据分析,制定更为优化的系统方案,使车辆行驶在最佳路径上,缩小行车时间、节省燃料、减少环境污染,提高路网通行能力和服务质量。在教育领域,利用大数据分析学生和学习习惯,改进课程、教学和学生学习方式,提高教学质量。

因而,大数据是城市智慧化的助推器。大数据在智慧城市各领域的应用将推动企业快速发展、产业转型升级、经济可持续增长,提高政府治理能力,提升公共服务水平。

7.1.2 大数据在国内外的应用现状

在国外,美国、英国、日本等发达国家都已开启了大数据战略的大幕,积极推动数据资源开放和利用,投入巨资进行大数据研究和应用,合力推进大数据产业化、市场化进程,大数据应用发展迅猛。国内智慧城市建设逐渐从概念走向落地,各地政府大力发展信息基础设施,推动政府和企业数据资源开放和共享,建设公共服务平台推动智慧应用。这些措施都为大数据在智慧城市领域的应用奠定了坚实基础,提供了有效的支撑。

1) 发达国家推行大数据战略,大数据应用发展迅猛

现在数据资源被认为是与自然资源、人力资源一样的战略性资源,大数据是当代最受关注的信息技术之一,被认为是人们获得新的认知、创造新价值的源泉和改变市场、组织政府与公民关系的方法。发达国家积极推动数据开放、共享,多个国家把大数据上升为国家意志,大幅提高大数据研究和应用投资,大数据应用迅猛发展。

(1) 支持数据开放,推动建立数据平台。各国政府是推动数据公开的主要力量,美国、欧盟等发达国家和地区积极推动政府数据信息的公开,建立国家层面的数据开放平台,服务政府、企业和公众。日本等亚洲国家也加快脚步建设数据开发平台和数据中心,促进数据开放。

美国具有重视数据的传统。自20世纪以来出台了一系列法规,对数据的收集、发布、使用和管理作出了规范,经过不断的发展和完善,如今美国数据、信用和隐私已形成较为成熟的法律框架和道德规范。2012年5月,美国数字政府战略发布,提出要通过协调化的方式,以信息和客户为中心,改变联邦政府工作方式,为美国民众提供更优的公共服务。其中关键就是政府组建平台或设备供民众随时随地获取政府信息和公共服务。美国政府的大数据平台 Data.gov^[1]作为数字政府的关键组成部分,已经开放了涵盖农业、气象、金融、就业、人口统计、教育、医疗、交通、能源等约50个门类的数据,并且提供应用程序、软件工具、手机应用插件,供公众和开发者对数据进行使用和开发应用。截至2012年11月,Data.gov平台共开放出了388 529项原始数据和地理数据。

英国政府早在2009年就开始打造在线数据公开系统。在全球范围内,英国政府的政务大数据公开“Open Data”项目最为成功^[2]。根据德勤的报告,2010年1月以来,英国政府 Open Data 网站(Data.gov.uk)的人均访问页面数增长了285%,总访问量比法国和美国的 Data.gov 的同类网站还要高。

法国政府投资部自从2006年以来支持了16个重大的数据中心项目。2011年7月,法国工业部宣布投资20万欧元,启动“Open Data Proxima Mobile”项目^[3],涉及交通、文化、旅游和环境等领域,该项目使所有法国公民以及在法国旅游的欧洲公民都能通过个人移动终端使用法国的公共数据。为了便于公民自由查询和下载公共数据,2011年12月,法国政府推出的公开信息线上共享平台Data. gouv. fr正式上线。

相对于美国和欧洲,亚洲数据开放起步晚。日本、韩国等出台一系列政策加快数据资源开发。2012年6月,日本IT战略本部发布电子政务开放数据战略草案,迈出了政府数据公开的关键性一步。2013年6月,安倍内阁正式公布了新IT战略——创建最尖端IT国家宣言^[4],提出尽快建立跨政府部门的信息检索网站,以便于企业利用政府的大量信息资源,计划到2015年末达到与其他发达国家同等的信息开放度。韩国政府宣布将建设一个面向中小型企业、风险企业、大学和普通公民的开放大数据中心,方便企业和民众通过该中心对大数据进行提炼和分析,利用大数据技术解决业务或者研究方面的问题。

(2) 投入巨资研究和应用大数据。为了在大数据领域抢占先机,各国将大数据上升到国家战略高度,发布一系列政策,在大数据研究和应用上投入了巨额资金。

美国是首个把大数据上升为国家战略的国家。2012年3月,奥巴马政府宣布了《大数据的研究和发展计划》,该计划中6个联邦政府的部门和机构计划新增2亿美元的投资,提高从大数据集中提取知识和观点的能力,加快科学工程的步伐,加强国家安全,并改变教学研究。美国国防部开展大规模数据集的异常检测和特征化、军事计算机网络与网络间谍活动检测、自动化和人机集成推理、大规模的军事图像分析等多个项目,加强大数据在国土安全方面的研究和应用。卫生和人类服务方面,医疗保险和医疗补助服务中心开发一个基于Hadoop的数据仓库,支持对医疗保险和医疗补助要求的分析和报告,引导和支持各种医疗保险和医疗补助服务中心高优先方案改进决策;国家档案和记录管理建立网络基础设施,对国家档案馆87万多样化的数字记录的文件和信息进行收集。在生物医药方面,美国国家卫生院及其下属机构建立多个生物医学资源信息中心和平台,提供包括医疗图像在内的各类信息资源共享和协作,促进医学数据资源重复利用和进一步开发,并为医学研究和临床实践提供决策支持。通过大数据战略的实施,美国在科研教学、环境保护、工程技术、国土安全、生物医药等多个领域实现了突破。

英国政府重视大数据研究和产业化。英国政府于2012年5月注资10万英镑,支持建立了世界上首个开放式数据研究所(the Open Data Institute, ODI)。英国政府通过利用和挖掘公开数据的商业潜力,为英国公共部门、学术机构等方面的创新发展提供“孵化环境”,同时为国家可持续发展政策提供进一步的帮助。英国经济在2008年金融风暴后持续疲软,在此情况下英国商业、创新和技能部仍在2013年1月宣布注资6亿英镑(约9.12亿美元)发展8类高新技术,其中接近总投资三分之一的1.89亿英镑用来发展大数据技术;2013年8月,英国政府发布《英国农业技术战略》,将高达6000万英镑农业技术投资将集中在大数据上,以促进英国的农业科技商业化。

2013年2月,法国政府发布《数字化路线图》,列出5项将会大力支持战略性高新技术,其中一项就是大数据。2013年4月,法国经济、财政和工业部宣布,将投入1150万欧元用于支持7个未来投资项目,通过发展创新性解决方案,并将其用于实践,来促进法国在大数据领域的发展。

日本政府提出:“提升日本竞争力,大数据应用不可或缺”。2012年7月,日本总务省ICT基本战略委员会发布了《面向2020年的ICT综合战略》,提出“活跃在ICT领域的日本”的目标,重点关注大数据应用所需的社会化媒体等智能技术开发、传统产业IT创新、新医疗技术开发、缓解交通拥堵等公共领域应用等。日本的大数据战略,以务实的应用开发为主,尤其是在和能源、交通、医疗、农业等传统行业结合方面。

(3) 大数据应用发展迅猛。国外政府层面大范围的数据开放和大数据研究应用的巨额投资为大数据应用创造了良好的环境。发达国家相继在交通、能源、公共安全等城市管理和公共服务领域开展了大数据实践,大数据应用遍地开花,发展迅猛。

美国具有重视数据和应用的历史传统,包括IT基础设施的完善、硅谷精神和创新力量,以及一大批的领军企业,因此大数据在智慧城市方面的实践走在了世界前列。例如,纽约市利用大数据防火^[5]。纽约市有大约100万栋建筑物,每年有差不多3000栋会因为火灾损毁。对付火灾,防患于未然才是最根本的处理方法。纽约消防部门应用数据和分析手段改变了FDNY(纽约消防局)日常建筑物检查的方式,帮助城市341个消防单位更准确地定位有潜在火灾危险的建筑物。根据不同的影响因素,消防部门从通过各种渠道收集到的数据里,划分出了60个可能会产生火险的类别,其中包括区域居民平均收入、建筑物年龄、是否存在电气性能问题等。从这些数据里能计算出它们的火险概率,并依此由高到低进行安全性排查。通过60面算法给33万栋建筑物都标注了风险指数。当进行每周检查时,电脑就将会根据这一分数给出一个列表,提醒他们哪些建筑物应该优先注意。有了数据挖掘项目的辅助,高危建筑物可以得到最高重视。除此之外,建筑物的数据挖掘对象是随机抽取的,其中学校和图书馆会受到格外重视。波士顿也在采用一个类似的模式来分配警力,以防止盗窃案的发生。佛罗里达州迈阿密戴德县将数十种关键县政工作和迈阿密市紧密联系起来,帮助政府在制定治理水资源、减少交通拥堵和提升公共安全等方面决策时提供了更好的信息支撑,这些都是美国大数据在智慧城市各个领域应用的典型案例。

英国在积极拥抱大数据的实践过程中已取得初步效果。据Policy Exchange提交的一份名为《大数据机遇》的报告显示,英国政府通过高效使用公共大数据技术,每年可节省约330亿英镑,相当于英国每人每年节省约500英镑^[2]。通过数据使用,优化政府部门的日常运行和刺激公共机构的生产力,可以为英国政府节省130亿~220亿英镑;减少福利系统中的诈骗行为和错误数量将为英国政府节省10亿~30亿英镑;有效地追收逃税漏税将为英国政府节省20亿~80亿英镑。

法国在智慧城市中的大数据应用方面也投入了大量财力,包括法国电信、施耐德集团和达索集团等诸多法国知名企业都在旗下设立了专门从事智慧城市设计和研发的工作室

或实验室,在政府引导下积极投身智慧城市建设。例如,法国电信公司 2008 年在法国卡涅和格勒诺布尔两个城市尝试利用智能通信技术优化城市居民生活质量^[3]。法国电信公司在卡涅中心城区的实验项目安装了数百个各类感应器,用来监控、测量、控制城市环境,并记录各方面的环境数据。根据法国电信公司的测算,卡涅的路灯照明占整个城市能源消耗的 40%,而利用感应器,城市街道照明和维护成本可以减少 20%~30%。法国里昂市与 IBM 利用大数据技术合作开发能缓解道路拥堵的系统——决策支持系统优化器。该系统基于实时交通报告来侦测和预测拥堵。当交管人员发现某地即将发生交通拥堵,可以及时调整信号灯让车流以最高效率运行。该系统对于突发事件也很有用,例如帮助救护车尽快到达医院。而且随着运行时间的积累,这套系统还能够“学习”过去的成功处置方案,并运用到未来预测中。

根据日本著名的矢野经济研究所预测^[4],2020 年日本大数据市场规模有望超过 1 兆日元。矢野经济研究所的报告显示,2011 年日本的大数据相关行业的市场规模为 1 900 亿日元,2012 年约为 2 000 亿日元,同比大约增长 5%。该机构同时预测,到 2013 年以后,每年将增长 20%;照此计算,到 2015 年将达到 4 200 亿日元,2017 年达 6 300 亿日元,2020 年约为 1 兆 500 亿日元。从日本大数据投资方向上来看,将数据利用于产品销售及新产品开发的 BI 关联 IT 投资目前占其总额的半数以上。2015 年以前,企业内部的数据利用将占主导地位,而 2017 年以后,能源管理系统等维护社会基础设施的系统需求将会推动大数据市场的发展。另一方面,日本大数据市场在其国内 IT 市场中所占比重,2011 年为 1.7%。假定今后日本的 IT 投资额不变,在此情况下,到 2015 年其比重为 3.8%,2017 年占 5.7%。据此分析,到 2020 年,大数据市场在日本 IT 市场总额中将占 10%左右。据日本总务省公布的 2013 年版《信息通讯白皮书》估算,充分利用大数据将给日本带来每年 7.77 万亿日元(约合 777 亿美元)的经济效益。

总之,大数据正在成为国家竞争的前沿,以及产业竞争力和商业模式创新的源泉。在这次云计算与大数据的新变革中,中国与世界的距离最小,在很多领域甚至还有着创新与领先的可能。一方面,大数据技术以开源为主,迄今为止,尚未形成绝对技术垄断。即便是英特尔、IBM、甲骨文等 IT 巨头,也同样是集成了开源技术,和本公司原有产品更好地结合而已。另一方面,中国人口和经济规模,决定中国的数据资产规模将冠于全球,客观上为大数据技术的发展,提供了广阔的发展空间。从国家角度来看,大数据是重要战略资源。因此,挖掘大数据价值,推动大数据发展,重点需要政府发挥作用。

2) 国内智慧城市建设落地,大数据应用刚刚起步

我国大数据在智慧城市建设中的应用,已在国家的引导下进入统筹推进阶段。2013 年是我国智慧城市建设从概念逐步落地的一年,现阶段智慧城市以加强信息化基础建设为主,大数据在城市各领域的应用处于研究和摸索阶段。

(1) 我国智慧城市建设进入统筹推进阶段。到 2013 年 8 月,我国试点的国家智慧城市已达到 193 个,在试点示范工作取得一定经验后,国家还将逐步鼓励和支持有条件的地方因

地制宜地推进智慧城市建设。为避免盲目无序建设,国家发布了一系列政策指导和规范智慧城市建设。

2013年7月,由8个部委共同起草的《关于促进智慧城市健康发展的指导意见》报请国务院,该指导意见提出了建成基础设施更加智能、公共服务更加便捷、社会管理更加精细、生态环境更加宜居、产业体系更加优化的智慧城市,并建议各地启动交通、电网、水务、环保、医疗、养老、社区、家居、教育、国土等10个领域智慧工程建设,深化重点领域的智慧化应用,为公众提供更加便捷、高效、低成本的社会服务。

2014年3月发布的《国家新型城镇化规划(2014—2020年)》,提出推进智慧城市建设要统筹城市发展的物质资源、信息资源和智力资源利用,推动物联网、云计算、大数据等新一代信息技术创新应用,实现与城市经济社会发展深度融合。强化信息网络、数据中心等信息基础设施建设。促进跨部门、跨行业、跨地区的政务信息共享和业务协同,强化信息资源社会化开发利用,推广智慧化信息应用和新型信息服务,促进城市规划管理信息化、基础设施智能化、公共服务便捷化、产业发展现代化、社会治理精细化。增强城市要害信息系统和关键信息资源的安全保障能力。

上述文件为各地智慧城市有序规范地建设指明了大方向,我国智慧城市建设进入统筹推进的新阶段。

(2) 各地提升信息基础设施,奠定大数据应用基础。现代城市本身是一个复杂开发的系统,智慧城市运用新一代全面感知技术更增加了城市数据量。传统IT架构无法支撑海量数据的收集、传送、分析和使用,更大带宽、无处不在的网络必不可少。

各城市已纷纷把信息网络基础设施提升纳入在智慧城市规划。北京在《智慧北京行动纲要》中提出了完善高速泛在的信息网络,建设“无线城市”,建设城乡一体的高性能光纤网络,推进“三网融合”,加快建设政务和公共服务高清视频传输网络;统筹建设全市政务信息基础设施,建设全市统一的传感终端网络、政务物联网数据专网、无线宽带专网及物联网安全保障体系,建设一流的数据中心,统筹建设全市便民服务终端网络,整合服务终端。天津提出2015年将基本实现城市出口带宽1.5TB以上,光纤入户覆盖率达到90%以上,无线宽带覆盖率达到100%;2020年城市出口带宽在2015年的基础上翻一番,光纤入户率达到98%以上的目标。上海在其推进智慧城市建设规划中提出了在2011~2013年构建国际水平的信息基础设施体系,推动实施宽带城市、无线城市、通信枢纽、三网融合、功能设施5个专项,落实完善规划体系、规范建设管理、强化机制建设3项重点任务。上海在2013年10月提前完成了光纤到户、无线宽带覆盖城镇化地区的目标。截至2013年10月,FTTH覆盖家庭数达到830万,新建住宅FTTH比例达到100%,新增3G/LTE基站2.4万个,重点场所3G覆盖率达100%,WLAN公共运营接入点数达到17.6万,3G移动电话用户达1074万户,固定宽带接入用户数达到617万,4MB以上用户占比达82.5%,其中8MB以上用户占比达66.6%^[6]。

目前,各省市的网络基础提升已取得阶段性成果,四网协同发展成效显著,宽带网络综

合能力大幅提升。公开资料显示,截至2012年年底,中国移动累计完成TD-SCDMA基站建设约28.3万个,WLAN接入点达到395万个,建设TD-LTE基站2.2万个,完成19个省、7541个行政村通宽带工作。铁通公司作为中国移动有线宽带运营的主体,宽带用户数超过1045万。中国电信2012年光纤到户覆盖家庭数新增3300万户^[7],达到6300万户,净增固定宽带用户1570万户,用户规模突破1亿户。中国联通2012年净增FTTH覆盖家庭1522万个,宽带端口超过1亿个,城市地区20M以上端口4330万个,占城市地区端口的53%;农村地区4M以上端口1740万个,占农村地区端口的65%^[8]。

随着各城市网络覆盖面加大和综合能力提升,城市信息化环境不断优化,为大数据在智慧城市的应用奠定了坚实的基础。

(3) 建设公共服务平台,支撑智慧城市各领域大数据应用。我国数据开放程度不高。从政府层面看,虽然各级政府发展电子政务,在政府网站积极地公开政务信息,但数据相对分散,存在信息孤岛现象。企业层面上,企业的数据大多储存在内部,不对外公开。工业企业的信息化程度不高,内部数据利用和分析相对较少,产生数据量大的电子信息、金融、能源等行业的企业数据分析和利用的热情高,相当一部分企业通过数据分析尝试精准营销和改进服务。

目前政府着手建立各类公共服务平台,促进跨部门、跨行业、跨地区的信息共享和业务协同,推动政务数据和行业数据的开放和应用。2013年11月20日,国家统计局与阿里巴巴、百度等11家公司携手共同建立统计局大数据应用平台。上海建立了食品安全监管和公共服务平台、社区事务综合服务平台、社区生活服务信息化平台等多个平台,为城市管理、公共服务领域智慧化提供了保障。城市各类公共服务平台的建立完善为大数据在智慧城市应用提供了有利条件,将有力支撑大数据在智慧城市各领域的应用。

(4) 城市各领域大数据应用逐渐增加。目前大数据技术在城市各领域的应用还在起步阶段。上海正在研究运用大数据缓解交通拥堵的问题,将通过建立一系列基于数据的交通管理系统,建立应对流量研判、道路疏导、楼宇规划、道路管制应急响应、特征车辆检索等一系列应对措施,在交通拥堵时疏导交通,提醒用户。在医疗领域,2013年6月,上海的医联网将覆盖全市所有层级的全部医院,医院的任何经营管理和诊疗业务将被纳入实时监控。除了这样普遍意义上的健康信息的收集和分析,以使技术在医疗诊断以及健康管理方面发挥优势以外,上海医疗卫生领域大数据推动将以强大的技术支撑着上海对公立医院的管理强化。2014年南京夫子庙灯会期间,大数据助力景区的信息分析,优化景区服务。南京市夫子庙—秦淮风光带管理办公室与中国移动南京分公司搭建的“客源分析系统”协助景区管理单位建立区域游客人口监测统计分析技术手段,完成当前游客流量统计功能,还可对用户开展大数据分析,根据游客群体来源、类型、特征等整体分布情况,为景区商铺提供营销建议,动态化显示景区经营效益,便于景区适时调整经营策略和优化服务。

总的来说,我国尽管已经意识到大数据的竞争力和战略性,却并未在真正意义上将其提升到国家战略高度,数据财富尚未形成全民意识。因此有必要从国家层面上制定大数据

发展规划,将大数据上升为国家战略。通过国家层面的战略规划明确大数据产业的发展重点、空间布局和保障措施,推动和改善与大数据相关的收集、储存和分析工具及技术,并在公共服务领域(如安防、医疗、卫生、教育等),开展大数据应用示范,提高应急处置能力和安全防范能力,提升服务能力和运作效率。

7.1.3 市场规模和发展趋势

政府和企业已经充分认识到了大数据提高社会生产力、提升创新能力的价值,对大数据的研究和应用保持了高度的热情。全球大数据仍会保持快速增长的态势,未来随着大数据技术的成熟,其投资重心将向大数据应用转移。随着我国智慧城市建设的推进,信息网络投资将持续升温,信息化水平将进一步提高,相应公共服务平台建成后,政府管理和公共服务中的服务型政府建设、社会保障、公共安全、市场监管、教育文化、医疗卫生、城市综合管理领域对大数据的需求会不断加大。

1) 全球大数据投资保持快速增长

麦肯锡全球研究院在 2011 年发布的大数据白皮书指出,大数据为美国的医疗服务业每年节省 3 000 亿美元,为欧洲的公共部门管理每年节省 2 500 亿欧元,为全球个人位置数据服务提供商贡献 1 000 亿美元;帮助美国零售业净利润增长 60%,帮助制造业在产品开发、组装等环节节省 50% 的成本。日本总务省公布的 2013 年版《信息通讯白皮书》估算^[9],充分利用“大数据”将给日本带来每年 7.77 万亿日元(约合 777 亿美元)的经济效益。“脸谱”、“网飞”大数据项目的成功,促使信息技术企业加快推动大数据业务的发展。越来越多的企业希望能够借助大数据方案控制成本、优化流程并降低管理风险。

因此,未来几年各国政府和企业对大数据投资的热情将持续升温,大数据全球市场规模仍会保持快速增长态势。IDC 预测 2017 年大数据市场规模将达 324 亿美元,年复合增长率为 27%。市场研究公司 Markets and Markets 预测,2013~2018 年,全球大数据市场的年复合增长率为 26%,预计将从 148.7 亿美元增长至 463.4 亿美元。日本著名的矢野经济研究所的报告显示,2011 年日本的大数据相关行业的市场规模为 1 900 亿日元,2012 年约为 2 000 亿日元,同比大约增长 5%。该机构预测到 2020 年日本大数据市场规模有望超过 1 兆日元^[4]。

中国作为大数据新兴市场,其市场规模增速将大大高于全球增速。IDC 预测中国大数据技术与服务市场规模未来 5 年的复合增长率将达 51.4%,从 2011 年的 7 760 万美元增长到 2016 年的 6.17 亿美元^[10]。

2) 投资重点将逐渐从研发和数据利用向行业应用转移

从大数据投资方向上来看,目前以技术产品研发和数据利用为主。根据 DJX VentureSource 对美国 2013 年风险投资的全年数据统计^[11],风险投资公司对数据库软件领域进行了 101 次投资,总额达 9.042 亿美元,高于 2012 年 95 次投资、1.818 3 亿美元的总

额。重要融资包括：数据中心基础设施解决方案提供商获投资 1.01 亿美元，云储存初创公司 Primary Data 和 Hortonworks 分别获得 5 000 万美元融资。风险投资公司在数据管理服务领域一年中共完成投资 103 次，总额达到 8.649 亿美元。日本矢野经济研究所的报告显示，日本把数据利用于产品销售及新产品开发的 BI 关联 IT 投资目前占其总额的半数以上。

随着大数据技术和产品的不断成熟，大数据领域的投资将逐步向应用领域过渡。日本矢野经济研究所预测，2015 年以前，本企业内部的数据利用将占主导地位，而 2017 年以后，能源管理系统等维护社会基础设施的系统需求将会推动大数据市场的发展。

3) 国内信息网络投资持续升温

各地智慧城市建设过程中，信息网络基础建设投资不断加大。已有多个城市携手移动、联通和电信三大运营商实施“宽带中国”计划，加快宽带网络建设的步伐。随着宽带中国计划的不断推进，信息网络建设投资持续升温。2013 年江苏省用于网络建设的总投资达 413.4 亿元，其中固定资产投资 262.8 亿元，直接用于“宽带江苏”、“无线江苏”工程建设投资达 184.2 亿元。江苏省通信管理局目前正全力推进江苏“2014 宽带中国专项行动”。青海投入专项资金支持宽带建设计划，确立总投资 113 亿元的 10 个大项、63 个小项的重点工程及项目。截至 2014 年 1 月底，有进展的项目为 36 个，完成投资 6.8 亿元，占年计划的 6%。

2014 年 2 月，上海市人民政府与中国联通集团有限公司在沪签署共建智慧城市战略合作框架协议，在 2014~2016 年进行战略合作，全面推进上海智慧城市新一轮建设。3 年内，中国联通将在上海投资 100 亿元，以信息化需求为导向，在着力推进重点区域建设的基础上，提升城市光纤宽带网络能级，构建第四代宽带无线移动通信网，规模推广互联网创新应用，规模推动智慧应用在各领域的融合。

在我国，数百城市争建智慧城市，金融机构投资支持建设力度加大。2013 年 1 月，国家开发银行与中国城市科学研究会签订“十二五”智慧城市建设战略合作协议^[12]，计划在“十二五”的后 3 年内，提供不低于 800 亿元的投融资额度，支持中国智慧城市建设。之后又有两家商业银行做出承诺，表示将提供不低于国家开发银行的授信额度，支持智慧城市建设，另有其他投资机构也在 4 月初签订了 2 000 亿元的投资额度，粗略估计，相关投资将超过 4 400 亿元。截至 2013 年 8 月，中国已有 320 多个城市及地区投入 3 000 多亿元建设智慧城市^[13]。据不完全统计，全国已有 95% 的副省级以上城市、76% 的地级以上城市，总计约 230 多个城市提出或在建智慧城市，计划投资规模近万亿元，到 2015 年将超过 2 万亿元^[14]。

4) 政府管理和公共服务领域大数据应用需求将会加大

我国多个城市在制定智慧城市规划时，将政府管理和公共服务的智能应用作为了重点。上海智慧城市规划提出重点推进城市建设管理、城市运行安全、智能交通、社会事业与公共服务、电子政务、信息资源开发利用、促进“四个中心”建设、“两化”深度融合 8 个专项，促进城市运行管理水平、经济发展水平、公共服务水平和居民生活质量明显提升。杭州计

划在交通、医疗、教育、公共服务、旅游、城管、环境、电网、水务、应急管理十大智慧应用系统建设取得突破性进展,并形成良好的服务运营模式。南京则重视建立智慧公共服务体系并在交通、医疗、教育、文化和社会保障等领域进行智慧应用以提升市民生活品质,提高居民幸福指数。对比了北京、上海、南京等地的智慧城市规划,下面几个领域是智慧城市中政府管理和公共服务需求的热点。

(1) 服务型政府建设:通过构建城市数据库和统一的服务平台,推进信息资源整合共享,为智慧城市应用体系建设提供全面、及时、准确的信息资源,为政府公共管理、企业生产经营和居民生活服务提供有力的信息支撑。推进智慧安居服务,创建智慧社区,为市民提供一站式信息服务。

(2) 社会保障领域的应用:深化人力资源、社会保障、社会救助领域信息化,建立健全覆盖城乡的社保信息服务体系,提高社会保障信息化水平,提升社保关系跨地区转接能力,促进劳动力的有序流动和社会充分就业。

(3) 公共安全领域的应用:通过进一步扩大城市视频采集及监控范围,建设完善覆盖全市应急指挥信息平台,显著增强城市防灾、减灾、救灾能力。强化公共安全信息化建设,加快推进社会治安综合治理信息平台建设,促进“平安城市”建设。建设事故隐患、重大危险源、危险化学品等数据库,提高安全生产监管的信息化水平,提升生产事故预防和处置的能力。

(4) 市场监管领域的应用:通过建立统一的市场监管综合执法信息平台,实现执法信息共享、业务协同和政务公开,可提高协同监管能力。加快建设产品质量与食品安全监管网络,提供准确、及时的信息服务。

(5) 教育文化领域的应用:在教育领域,推动在线教育,提高教育网络化应用基础资源大众化、促进教育优质资源充分利用和均衡配置;在文化领域,推进数字图书馆、数字博物馆等文化信息资源共享工程,发展新兴文化业态,加快公共文化信息资源开发和信息服务体系建设。

(6) 医疗卫生领域的应用:通过促进健康档案、电子病历的整合应用,推进区域卫生信息平台建设,优化网上挂号、双向转诊、远程医疗、疾控预警、药品监管公共医疗服务。

(7) 城市综合管理领域的应用:采用物联网、云计算、大数据等技术建设智能交通系统,提高城市交通管理、指挥和调度水平,提升交通运行效率,保障城市畅通有序。逐步整合交通、通信、水、电、气等城市生命线的运行监控信息,实现对城市运行的精确、综合监管,提高城市管理整体效率。

总之,在智慧城市建设不断推进的过程中,城市信息化水平会进一步提高。随着相应公共服务平台的建成,这些领域对大数据的需求会不断加大。

5) 信息企业布局转型提供大数据服务,挖掘巨大的商业利益

大数据时代,越来越多的企业开始认识到大数据对企业运营以及企业未来发展方向的重要性。微软、谷歌、英特尔、甲骨文、IBM等著名跨国企业和阿里巴巴、腾讯、百度、曙光、浪

潮、华为国内重量级信息企业都开始抢占大数据产业链的各个环节,加快推动大数据业务的发展,对于百度、阿里巴巴、腾讯等具有海量数据的互联网企业,通过投资合作收购扩大其数据资产的覆盖面,逐步向数据公司转型。与此同时,这些企业展开了和政府的一系列合作,加快大数据在智慧城市各领域应用的研究和实践,将大大推动智慧产业和智慧应用的发展。

以 IBM、甲骨文、曙光、华为为代表的企业通过投资不断加强其大数据各环节的技术水平,通过发布大数据集成系统产品向大数据业务公司转型。以 IBM 为例,该公司出资 160 亿美元收购了超过 30 家大数据企业,其中以数据挖掘和数据分析领域的收购居多,2013 年收购的星分析(Star Analytics)公司就是一家业务分析软件制造商。通过收购,将 IBM 产品与这些数据挖掘和分析公司的软件产品组合,完善 IBM 在性能管理和业务智能方面的解决方案。作为目前为数不多能够集咨询、服务、软硬件综合实力,提供端到端全面整合解决方案的厂商,IBM 从 2012 年开始,连续 3 年内将在中国投资超过 3 亿人民币,用于进行大数据的研发与推广。2012 年,IBM 在北京成立全球首个大数据智慧赋能中心,针对以中国为代表的新兴市场,调动 IBM 的全球资源,通过 IBM 中国开发中心,将 IBM 软件在大数据、信息管理、商业分析等相关领域的前瞻洞察和创新能力引入中国市场,帮助中国的政府和企业在大数据时代对于快速转型和智慧成长的需求。从 2013 年开始,IBM 在中国的大数据投资逐渐从北京、上海等一线城市向我国二级城市扩展。2013 年 8 月,IBM 牵手绵阳投资大数据,启动 IBM 大中华区首个大数据分析竞争力中心项目。与此同时,IBM 正展开对河南的密集布局。

埃森哲和甲骨文公司携手发布了一项战略性计划,该计划将持续多年,将 Oracle 集成系统更快速和更具成本优势地提供给客户。埃森哲创建了一整套解决方案,将 Oracle 集成系统整合到了埃森哲领先的数据中心转型咨询和外包服务中。这些解决方案可使客户利用其数据中心管理、处理和分析在数字化世界产生的大量数据,同时,帮助企业通过整合服务器、存储和网络组件来显著降低数据中心的总体拥有成本。包括 Oracle Exadata 数据库云服务器、Oracle Exalogic 中间件云服务器、Oracle SuperCluster 和 Oracle Exalytics 商务智能云服务器在内的 Oracle 集成系统,通过预先集成硬件和软件,可降低 IT 架构的成本和复杂性,并提高生产力和性能。利用全面的 Oracle 硬件和软件组合产品的广泛性能,埃森哲为 Oracle 集成系统建立了一系列补充。埃森哲和甲骨文公司也在投入、执行和实施基于 Oracle 集成系统的转型行业解决方案。例如埃森哲生命科学云能帮助客户对大量的医疗数据进行分析,同时提高药品全面的开发和增强功能。埃森哲公共服务平台是一个私有云平台,可为医疗改革加快系统部署。针对全球企业对更好的管理流程和更好的可视性的需求,埃森哲部署了云和集成中心架构,并部署了 Oracle 电子商务套件和 Oracle Exadata 数据库云服务器。埃森哲采用了关键技能和资源来快速地提供业界领先的云平台,利用 Oracle Exadata 数据库服务器来指导企业快速推动全球业务发展。

曙光公司是国内拥有众多云计算中心、行业云、城市云的公司之一,在 2013 年 5 月对外

发布了其大数据战略——“平台一体,智汇应用”,并推出了曙光大数据的核心技术和产品: XData 大数据一体机、XData-Hadoop 大数据处理软件等自主可控的创新产品,从数据收集、数据存储、数据分析、数据应用等方面构建了完整的大数据生态环境,携手政府、金融、电信、军工、教育开拓实践,在大数据“元年”共享硕果。这是曙光通过发展大数据从硬件设备供应商向解决方案和服务提供商转变的重要一步。曙光的大数据服务内容,更多的是针对不同的行业客户提供定制式的大数据解决方案。第一步帮助客户进行“数据落地”,即先把数据高效、低成本地存起来;第二步是“分析简化”,化繁为简,基于当前的数据处理、数据标准化的软件系统,进行一个分析、简化、编制,让它更有序;第三步是“价值新生”,即应用开发,根据顾客的需求,为它量身定制非常适合这个行业应用的一体化系统,实现业务集成化和决策智能化,而中国银联、中国移动、中国电信、中国农业银行、上海同济大学等早已是曙光大数据服务的客户,此外,在交通、医疗等领域,曙光也已有诸多实际案例。

阿里巴巴多年来积极打造数据生态圈,推动公司从电子商务企业向数据公司转型。2009年,阿里巴巴成立阿里云打造以数据为中心的先进云计算服务平台阿里集团;2012年,设立首席数据官岗位(CDO),推动“数据分享平台”战略推进;通过一系列入股、合作、并购动作,新浪微博、陌陌带来的社交数据,高德带来的地理数据,UC浏览器带来的移动浏览数据,虾米带来的音乐数据,快递打车带来的交通数据,余额宝带来的金融数据,与美的、海尔、创维合作带来的智能家电数据,还有墨迹天气、友盟、美团等几乎涵盖了人们生活的方方面面。再加上阿里巴巴原本拥有交易数据和信用数据,一个较为完整的数据生态圈逐步显现。阿里巴巴投资触角延伸到涉及生活的各个角落。在2006年10月收购口碑网进军分类信息领域,之后又相继投资了美团网、丁丁网等生活平台入口。2013年5月,阿里巴巴以2.94亿美元购买高德软件28%股份。2014年2月,阿里再次以10.45亿美元现金收购高德软件余下的股份。阿里巴巴通过高德地图建立国内最大的POI数据库后,将为O2O奠定坚实的数据基础。阿里巴巴集团计划通过基于地理位置的精准推送,把身边优惠和附近的活动推送给用户。而未来基于LBS的地理位置的移动服务,从智能出行、智能医疗到智能教育,覆盖社会民生各行业的方方面面。阿里快的打车与腾讯滴滴打车补贴之争和支付宝二维码支付是阿里巴巴推动移动支付的表现。阿里巴巴通过与家电企业的投资合作展开在智能家居领域布局。2014年,阿里巴巴与美的合作,通过收集消费者行为数据帮助美的对数据深度挖掘,构建新型制造业商业模式。这不是阿里巴巴第一次与家电结缘,2013年9月,阿里巴巴曾联合创维数码推出互联网电视。2013年12月,阿里巴巴集团也曾向海尔集团子公司海尔电器注资28.22亿港元。此外,阿里巴巴利用其在数据和平台上的经验和优势,与政府合作推动数据共享开放,发展大数据应用。阿里巴巴、百度等11家公司与国家统计局携手共同建立统计局大数据应用平台,共同研究探讨建立大数据应用统计标准,共同研究确定利用企业数据完善、补充政府统计数据的内容、形式和实施步骤等。2013年11月,阿里巴巴同河南省政府围绕菜鸟网络、智能骨干网、电子商务物流配送等签署一揽子合作协议。

百度方面,现在的数据总量接近 1 000 PB 左右,网页的数量大概是几千亿个,从大量的网页里面拿出几百亿来建索引,对于大量的网页,百度每年更新几十亿左右,每天享受用户的查询也会在几十亿左右。根据公开的信息,作为全球最大的中文搜索引擎,百度每天响应来自 138 个国家和地区的数十亿次请求,百度每日新增数据 10 TB,要处理超过 100 PB 的数据,从浩如烟海的信息中精确抓取约 10 亿网页,同时索引库还拥有千亿级在线索引能力,以帮助用户完成搜索过程。百度正在通过自建数据中心,开发了自己的大数据存储系统,并使用了多项新技术。计划投资 47.08 亿元的百度云计算(阳泉)中心已于 2012 年 8 月奠基,该项目预计 2015 年完工。建成后的百度云计算(阳泉)中心数据存储量将超过 4 000 PB,可存储的信息量相当于 20 多万个中国国家图书馆的藏书总量。多媒体部门是去年百度新成立的部门,主要从事语音、图像等方面的研究,通过深度学习模型,找出数据的价值,这些都是未来大数据重要应用,而语音、图像和视频都是非结构化数据,处理和分析难度更高,但非结构化数据将是未来大数据的主流。李彦宏成立了深度学习研究院(Institute of Deep Learning)并亲自任院长,还亲自出马去招揽人才,而这一切目的是通过大数据“深度学习”出互联网的未来。

腾讯方面,在 2013 年大数据大会上,在国内互联网体系中,腾讯拥有的数据覆盖多个领域,总存储数据量经压缩处理以后在 100 PB 左右^[15]。腾讯 QQ 目前拥有 8 亿用户,包括 4 亿移动用户,在数据仓库存储的数据占用单机数量已达到 4 400 台,总存储数据量经压缩处理以后在 100 PB 左右,并且这一数据还在以日新增 200~300 TB、月增加 10% 的数据量不断增长,现在正在为 1 000 PB 做准备。作为亚洲最大的数据承载中心,腾讯天津研发与数据存储中心正在滨海新区紧张地建设中,腾讯天津中心项目总建筑面积 9.33 万 m²,预计服务器托管能力超过 10 万台。另外,腾讯宣布斥资 2.15 亿美元收购京东商城 15% 的股权,并在后者上市时追加 5% 的股权。

总之,数据正成为与物质资产和人力资本相提并论的重要生产要素,大数据的使用将成为未来提高竞争力的关键要素,大数据在商业领域的应用,可以通过对顾客群体细分来实现对每个群体量体裁衣般的独特的行动,通过模拟现实来发掘新的需求和提高投入的回报率,通过分析消费者的行为特征进行商业模式、产品和服务的创新等,这些应用将会产生巨大的商业价值。

7.2 业务模式

根据维基百科的定义,“大数据”指无法在一定时间内用通常的软件工具进行捕获、管理的数据集合。基于此,大数据业务可被定义为:以新数据处理技术为手段,在海量、结构复杂、内容多样的数据集中,以较快速度解析出规律性的预见、趋势或判断。

美国互联网数据中心指出,互联网上的数据每年将增长 50%,每两年便将翻一番。在中国,大数据也会有比较大的发展空间,据估算,未来中国大数据潜在市场规模有望近 2 万亿元。总体而言,大数据技术及业务刚刚起步,展望未来,一片蓝海,因此深入挖掘大数据业务的商业模式必将成为企业的竞争战略之一。

本节首先介绍了大数据业务的产生背景和商业模式的涵义,然后总结了大数据业务发展的 6 种商业模式,并对每一种商业模式进行了定义、案例分析、优势归纳和关键成功因素探究,最后提出了大数据业务运营中存在的一些问题及对策。

7.2.1 商业模式的涵义

著名管理学大师彼得·德鲁克曾说过,当今企业间的竞争,不是产品的竞争,而是商业模式的竞争。Rappa(2004)认为,商业模式规定了公司在价值链中的位置,指导着公司如何赚取剩余价值;并指出商业模式明确了一个公司开展什么活动来创造价值,在价值链中如何选取上下游合作伙伴以及怎样与客户达成交易、为客户提供价值。

商业模式即企业通过产品或服务在价值链上下游主体之间建立的一种商务关系,包括公司所能为客户提供的价值、公司的内部组织结构、合作伙伴关系网络等用以实现这一价值并产生可持续盈利收入的要素。而大数据业务的商业模式就是围绕大数据资产和技术衍生出来的商业模式。

由于大数据背后的动因、参与主体、方法论的多样性,其秩序的建立也必然是一个循序渐进的过程,无法一蹴而就。这种秩序既涉及方法论和数据模型等技术层面的东西,也涉及商业模式和产业形态这样的上层建筑。无论是对行业应用的纵深还是通用技术的研究,无论是对初创公司还是产业巨头,抑或是企业内部的研发项目,大数据的招数并无定式。

在美国著名科技网站 Venturebeat 盘点的“大数据”界特征鲜明的 10 家初创企业所研究的课题,既包含社交化数据反馈、数据可视化这样的通用模型,也包含面向企业级的 Hadoop 应用、商业智能这样的方法论,还包含面向电商、零售商市场营销应用这样的行业纵深。但在看似无序的发展状态下,IT 巨头们围绕大数据产业链进行的收购却日渐增多,这或许就是产业秩序建立的开始。

7.2.2 大数据业务的商业模式类型

大数据产业链自底向上主要由三层构成。第一层是企业内部交易数据和企业外部的用户行为数据、物联网数据等,这一层次的主要任务是数据的采集、存储和传输等工作;第二层次是信息层,去粗取精,提炼后形成价值密度更高的信息,这一层可以产生诸如数据包销售、租赁等业务模式,也会诞生一批靠搜集各类数据为主业的公司,如区域数据提供商;第三个层次是知识层,对于知识的利用需要人工介入以外,主要还需要融合行业信息。具

体来看,围绕上述三个层次衍生出6种主要的业务模式^[16,17]。

1) 租售数据模式

将产业定位在大数据采集和整理阶段,通过收集、整理、过滤、校对、打包、发布等一系列流程后,实现数据的增值,这就是租售数据模式。作为中国领先的导航地图、动态交通信息及汽车综合信息服务提供商,四维图新致力于为全球客户提供专业化、高品质的电子地图数据产品和服务。其拥有全国最大的高质量导航电子地图数据库,建成了以北京为中心、覆盖全国的本地化导航电子地图数据采集更新体系,在基于静态的地图数据基础上不断加入实时动态的交通信息、丰富的生活信息和全面的地理信息。

租售数据模式对于数据提供商来说具有极大的价值,因为这一模式能使其拥有很强的话语权。由于数据的稀缺性,数据提供商位于产业链的有利位置,具有较强的议价能力、较强的竞争优势以及良好的成长空间。

这一模式的关键成功因素是大数据的采集和维护,企业要将在经营中接触到的大量实时数据进行汇总记录并校对,加工成客户所需的数据才能销售获利。

2) 租售信息模式

将产业定位在大数据整理和分析阶段,采编各类信息、数据,建设和维护数据平台,并通过各类渠道将信息传递、推广、销售出去,这就是租售信息模式。成立于1982年的美国彭博资讯公司(Bloomberg)是目前全球最大的财经资讯公司,其仅用了22年的时间,就将它的金融数据市场的销售收入超越了具有150年历史的、世界上最大的资讯公司——路透集团。彭博是全球商业、金融信息和财经资讯的领先提供商,通过其强大的信息、专家和咨询网络为全球重要的决策制定者带来关键信息。彭博的优势在于通过创新的技术来快速、精准地传递数据、资讯和分析工具。

租售信息模式能够成为企业竞争的法宝,企业结合终端业务比竞争对手更及时、更客观地提供相关信息和资讯给广大用户,可以抢占更多的市场份额。

这一模式的关键成功因素是采编各类信息资讯,要做到这一点,企业应建设和维护大型数据平台,并协同多种渠道进行信息和资讯的推广。

3) 数字媒体模式

将产业定位于媒体上,利用数据挖掘技术帮助客户开拓精准营销,企业收入来自客户增值部分的分成,这就是数字媒体模式。这类企业成长非常快,一般擅长数据挖掘分析技术,帮助一些数据大户如银行、运营商开展新的业务。亿赞普(北京)科技有限公司是一个高科技公司,基于技术和商业模式创新,搭建了全球化的云媒体平台。亿赞普目前已拥有56项国际核心专利,尤其在大数据处理和数据分类技术上处于国际领先地位。亿赞普云媒体平台创新的商业模式,包含了电信业、媒体、电子商务、广告服务等行业,构建了一条全新的数字媒体服务产业链,将助推互联网媒体产业链转型升级。

传统的互联网营销是完全碎片化的,广告主每一次广告投放的数据难以进行关联和复用,导致了广告费用的浪费。而数字媒体模式基于领先的数据挖掘技术,解决了这一难题,

不仅实现了跨媒体的广告调度,而且帮助广告主不断积累和复用自己的营销数据库,实现了营销活动的持续性和科学管理。

这一模式的关键成功因素是基于大数据分析和挖掘而积累的互联网知识。而其基于知识模式的经济价值和社会价值还远远没有发掘出来,其发展空间不可估量。

4) 数据使能模式

将产业定位在某一具体行业,通过大量数据支持,对数据进行挖掘分析后预测相关主体的行为,以开展业务,这就是数据使能模式。最典型的是小额信贷公司,在大数据时代,评估这些小微企业甚至个人还款能力的技术手段有了巨大进步,通过分析这些企业往来的交易数据、信用数据、客户评价数据等,完全可以掌握他们需要的资金量,甚至可以测算他们可能的还款时间,放贷风险大为降低。目前基于数据分析的小额信贷公司如雨后春笋,国内代表性的公司是阿里巴巴旗下的阿里巴巴金融。阿里巴巴金融承担阿里巴巴集团为小微企业和网商个人创业者提供互联网化、批量化、数据化金融服务的使命。其通过互联网数据化运营模式,为阿里巴巴、淘宝网、天猫网等电子商务平台上的小微企业、个人创业者提供可持续性的电子商务金融服务,向这些无法在传统金融渠道获得贷款的弱势群体提供“金额小、期限短、随借随还”的纯信用小额贷款服务。

数据使能模式依据大数据技术开展高收益、低风险的业务,为企业创造新的盈利模式。未来将会有更多的数据使能型的业务模式出现,它们将具备创新业务的特质。

这一模式的关键成功因素是维护数据的真实性和完整性,并适时进行风险分析。数据越完善,风险越低,越有利于保证企业的高收益。

5) 数据空间出租模式

将产业定位于大数据计算基础设施上,通过出租一个虚拟空间,从简单的文件存储,逐步扩展到数据聚合平台,这就是数据空间出租模式。Dropbox 是一个网络存储服务、网络备份工具和文件同步工具。其在线存储服务通过云计算实现互联网上的文件同步,用户可以存储并共享文件。用户可以通过 Dropbox 桌面应用软件,把档案放入指定文件夹,然后档案就会被同步到云端,只要用户在其他设备上登录自己的 Dropbox 客户端,都可以访问和管理自己 Dropbox 上的文件。

数据空间出租模式给个人和企业用户提供了实用的文件同步、备份、共享工具。另外,也可以很方便地分享给其他人。而自动备份的功能则大大提高了文件的安全性。这一模式的关键成功因素是平台的开发和维护,因为这一模式普遍的运作方式是后台自动备份指定的文件夹内容到云空间上,所以往往需要一个功能十分强大的开发平台来支撑。

6) 大数据技术提供商模式

将产业定位于大数据技术和工具上,围绕 Hadoop 架构开展一系列产品研发、技术服务,或是开发非结构化数据处理技术,这就是大数据技术提供商模式。狭义的大数据技术相关公司围绕 Hadoop 技术,提供大数据存储、检索、数据挖掘等应用。广义而言,大数据的核心技术之一是非结构化数据的处理技术,包括语音、视频、文本、图片等。拓尔思

是国内非结构化信息处理的龙头企业,公司专注于海量非结构化信息处理为核心的软件研发、销售和技术服务,其大数据管理系统 V7.0 兼容 Hadoop 标准,支持 PB 级海量数据管理。

大数据技术提供商模式迎合了大数据时代对海量数据进行挖掘整合的需求,而且移动互联网时代的海量消费数据给其发展带来了巨大的市场空间和成长机会。

这一模式的关键成功因素是准确把握技术发展方向,并保证提供优质的技术服务。同时,公司应构建清晰的营销网络架构,并且针对不同客户群体提供差异化服务,保证满足重点客户的定制化需求。

7.2.3 大数据业务运营中的问题及对策

总体而言,大数据技术及业务刚刚起步,前景一片光明,但其中也隐含一些“礁石”,应引起注意,概括起来有如下几点:

1) 数据挖掘的隐私问题

数据隐私、数据安全问题是大数据时代不可避免的问题。现阶段消费者已经生成了很多数据记录,比如电话记录、上网痕迹、交易记录等,凡使用数字化工具都会留下记录。将来这些数据在各个环节打通以后,一个人的行为就无法隐藏,数据持有人可以从历史数据中推测判断出行为人的行为轨迹和思维轨迹。人们往往强调数据整合能实现“ $1+1>2$ ”的效应,却忽略了这种扩大的效应对隐私的渗透力也是大于 2 的。我们缺少“数据权”的体现之一正是层出不穷的个人数据泄露事件。

由来自多个国家、多个组织的代表提出的“隐私偏好平台计划”是一种针对隐私泄露的技术保护策略,它是万维网集团在线隐私国际标准协议的一部分。计划旨在提供便于用户控制个人信息的一系列工具和服务,从而提高 Web 服务商和个体用户之间的相互信任。使网站以一种能被客户机分析的标准化结构表达网站的隐私政策,尤其可以使网站拥有者把隐私操作翻译成基于 XML 的计划声明,这种计划声明可被浏览器自动地检索和翻译。这意味着用户可以轻易地发现和理解特定站点的政策,并可作出是否接触该网站的明确决定。这一计划为解决数据挖掘中的隐私问题提供了极大的帮助。

2) 数据驱动的创新问题

不可否认,大数据时代日渐发展的数据驱动式的决策模式降低了决策风险,但也扼杀了潜在的、没有数据支持的创新。在这种情况下,人们的创新思维和创造力会逐渐被埋没,像乔布斯这样不相信市场调研,依靠本能和悟性创造了 iPhone 奇迹的天才会越来越少。

为了避免因数据驱动阻碍创新,企业应该引进和培养创新型人才,同时还须打破“封闭创新”,建立多元主体协同互动的创新模式,即企业应主动寻求与高校、科研机构、上下游等外部创新组织的技术合作,建立由企业主导的产学研合作市场机制,弥合断裂的创新链和

产业链,这样才能在智慧经济时代的全球竞争中脱颖而出。

3) 数据甄别的质量问题

数据甄别过程中的质量问题也是不可忽略的。面对如潮水般的数据,如果不加以筛选、甄别,就难以保证数据的完整性与客观性,在此基础上的数据分析与整合必然也会错漏百出,失去了其使用价值。肯尼斯·库可耶等编著的《大数据时代》一书在肯定了大数据的核心功能是预测之外,也指出大数据的不利影响是滥用大数据进行预测。

数据质量评估和监控是解决数据质量问题的重要手段,一般认为数据质量是一个层次分类的概念,每个质量类最终分解成具体的数据质量维度,如准确性、完整性、一致性、最小性等。数据质量评估的核心在于具体地评估各个维度,目前方法主要分成两类:定性策略和定量策略。其次,在评估后还需要提高数据质量。数据清洗(Data Cleaning)是数据质量提高技术研究的主要内容,主要集中在几个方面:重复对象检测、缺失数据处理、异常数据检测、逻辑错误检测、不一致数据处理等。

本节重点总结了大数据业务发展的6种商业模式,并对每一种商业模式进行了定义、案例分析、优势归纳和关键成功因素探究。随着社会经济的发展,未来还会产生许多新的商业模式。

尽管大数据面临一些小问题,但仍是蓬勃发展的趋势,大数据技术是数字化、信息化和智能化发展的必然,它为人类全面、深刻地认识世界和认识自身提供了新的方式和视角,这在信息时代以前是无法办到的。随着世界各国大力发展大数据业务,大数据技术必将为人类文明的发展作出巨大贡献,并成为驱动社会发展的巨大动力。

7.3 指标体系

7.3.1 概述

2007年10月,以维也纳理工大学 Rudolf Gimnger 教授为首的研究小组,从智慧人群、智慧经济、智慧治理、智慧流动、智慧环境、智慧生活等6大层面出发,构建了包含31项二级指标、74项三级指标的智慧城市评价体系,进而在对指标体系进行标准化变换与加总后,对70个欧洲中等规模智慧城市的发展水平进行了测算与排名。结果表明,瑞典、芬兰、荷兰、比利时、卢森堡、奥地利等国家的城市智慧程度较高^[18]。

IBM公司于2009年8月发布了《智慧的城市在中国》白皮书,提出智慧城市建设应该基于人(公共安全、医疗教育与生活质量)、商业(商业计划、对外开放、投资、劳工立法、产品市场立法等)、运输(公共交通网络、海运和空运)、通信(电子通信的基础架构,如电话、宽带和无线网络)、水(水的循环、供应与清洁)和能源(生产、运输体系与废弃物处理)等6大核心

系统,并指出这些系统的有效性、高效性和安全性决定一个城市如何运作和实现自身目标,以获得城市发展上的成功。6大系统的划分为构建智慧城市一级评价指标体系提供了有益参考^[18]。

国际智慧城市组织(ICF)作为一个长期关注智慧城市发展的智囊团。以智慧社区建设为核心进行智慧城市的研究,寻求智慧城市的最佳实践,以期推动城市的可持续发展。ICF主要从宽带连接、知识型劳动力、创新、数字融合、社区营销与宣传等5个方面去评价智慧社区的发展水平,并于2012年1月公布了“2012全球顶尖7大智慧社区”入选名单,美国奥斯汀、芬兰奥卢、加拿大魁北克、美国滨江、加拿大圣约翰、加拿大斯特拉特福与中国台湾台中市并列全球顶尖7大智慧城市^[18]。

2011年5月,上海浦东新区率先对智慧城市概念进行系统梳理,并提出了“智慧城市评价指标体系”。该指标体系统筹考虑了城市信息化水平、综合竞争力、绿色低碳、人文科技等方面因素,形成了45项具体的量化指标,其中15项为核心指标、30项为一般指标。这45项指标除了城市基础网络设施等硬件建设外,还包括智慧化交通管理、医疗教育体系、环保网络、社区管理,产业可持续发展能力、市民文化科学素养等软件条件。如果一座城市的两项指标达标率均超过80%,那么它将迈入“初级智慧型城市”的门槛;如果核心指标实现100%达标,一般指标的达标率也超过90%,那么它将成为“成熟智慧型城市”^[19]。2012年4月,中国工程院确立的“中国智慧城市”试点城市名单出炉,最近又提出“管理与服务水平、产业与经济水平、城市建设与环境水平、信息化水平和居民素养水平”是智慧城市评价国际标准体系的5项核心指标。2012年12月,住建部印发了《国家智慧城市试点暂行管理办法》和《国家智慧城市(区、镇)试点指标体系(试行)》,试点指标体系包括保障体系与基础设施、智慧建设与宜居、智慧管理与服务、智慧产业与经济四个一级指标。其中,保障体系与基础设施指标分为保障体系、网络基础设施、公共平台与数据库三类二级指标,涉及无线网络覆盖、城市公共基础数据库和信息安全等多项三级指标。智慧建设与宜居指标分为城市建设管理、城市功能提升两个二级指标,涉及数字化城市管理建筑节能、垃圾分类与处理、供水排水燃气系统等多项三级指标。智慧产业与经济指标包括产业规划、产业升级和新兴产业发展三方面,涉及创新投入、产业要素聚集、传统产业改造和高新技术产业等内容。

智慧城市建设的标准涉及基础性标准、技术标准,以及建设、运行、治理、工程管理标准和安全与行业应用标准。其中技术标准涉及云计算、移动互联、大数据、新兴技术应用等标准。因为智慧城市涉及面非常宽泛,需要应用很多技术,并开展更多信息、通信、应用领域等的协作。在标准的建立方面,急需国家出台相关规划或者指导建议。智慧城市建设目前已经有发改委、工信部、科技部、住建部等多个部委在布局,如何在国家层面形成统一科学完善的标准评价体系,是一个紧迫的课题。

本节试图从学术角度,整合国内外关于智慧城市指标体系已有的研究成果,并结合中国工程院汪应洛院士承担的中国工程院“关于智慧城市指标体系研究”的成果,来设计

智慧城市建设和评价的指标体系,以期为制定国家标准和地方智慧城市建设提供学术参考。

7.3.2 评价原则

1) 科学性原则

指标体系要客观地反映系统发展的内涵、各子系统和指标之间的相互联系,并能较好地度量区域可持续发展目标的实现程度。

2) 层次性原则

衡量社会的发展行为与发展状况是否具有可持续性,应在不同层次上采用不同的指标。

3) 相关性原则

可持续发展实质上要求在任何一时期,经济的发展水平或自然资源的消耗水平、环境质量和环境成灾状况以及人类的社会组织形式之间处于协调状态。

4) 明性原则

指标体系中的指标内容简单明了,具有较强的可比性。指标不同于统计数据和监测数据,必须经过加工和处理,使之能够清晰、明了地反映问题。

5) 可操作性原则

所确定的各项指标,其参数易于获取,整体指标体系有很强的实用价值。

7.3.3 评价指标

智慧城市的建成犹如一棵参天大树的成长一样,离不开大树所处的土壤、树本身的特点,所表现出的枝繁叶茂虽有共同点,却又处处体现差异。为此,可以用智慧之壤、智慧之干、智慧之花形象地描述智能城市的指标体系,它是一个综合性的评价指标体系,其一级指标有9个,分别为:资源禀赋、智慧基础设施、智慧平台、智慧人才、智慧治理、智慧环境、智慧社会、智慧民生和智慧经济,如图7-1所示。自然禀赋是智慧城市建设的先天条件,智慧基础设施、智慧平台、智慧人才、智慧治理、智慧环境是智慧城市建设的核心要素,智慧社会、智慧民生和智慧经济是智慧城市的直接表现,这些因素相互影响、相互促进。

值得注意的是,不同城市由于其自然禀赋、城市发展战略、城市定位及其自身人口、资源和环境特点等多方面的差异,因此又可以细分为:智慧旅游城市、智慧创新城市、智慧制造城市、智慧商务城市、智慧金融城市等,这就要求首先应该构建全面的、适合评价大城市、涵盖各个方面的综合性评价指标,综合性的智慧城市指标应力求全面、详尽,以真实刻画城市智慧程度。依据综合性指标体系,抽取、裁减具体评价指标以构建适合于评价不同特色

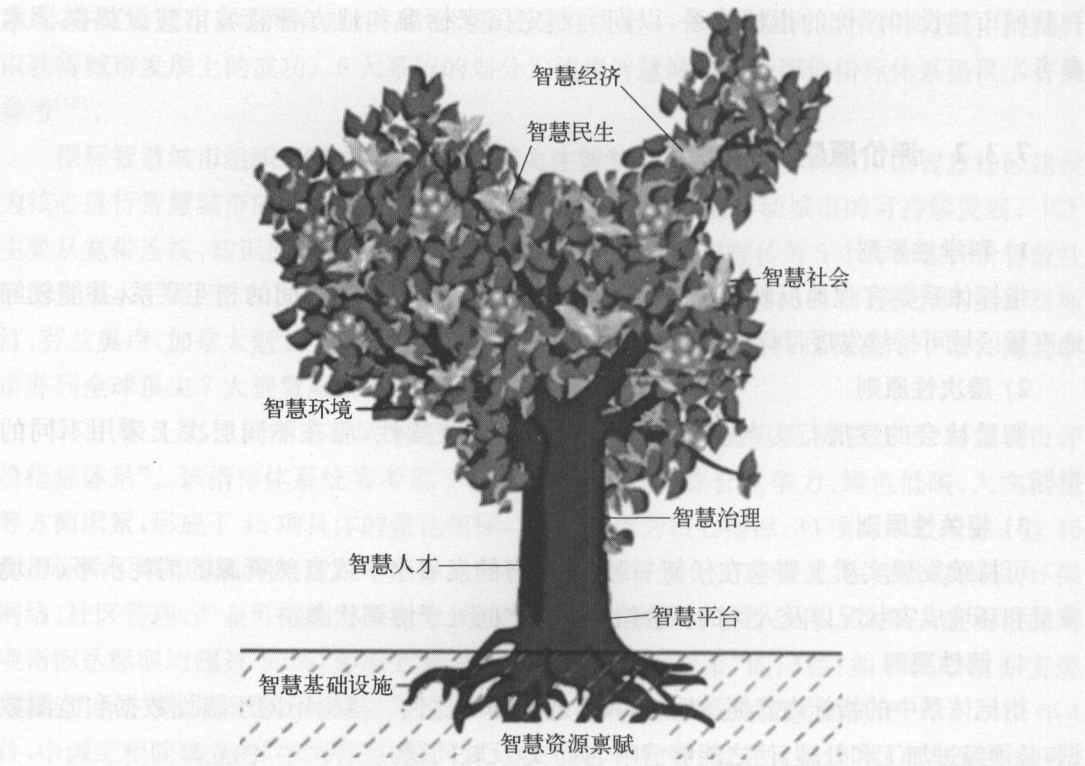


图 7-1 智能城市评价指标体系

城市的板块性指标体系,以刻画不同类别城市的智慧程度。综合性的、理念性的智慧城市的指标体系如表 7-1 所示。

表 7-1 智慧城市的指标体系

属 性	9 个 一级指标	22 个 二级指标	99 个 三级指标
智能之花	智慧经济	传统升级 (智能化)	传统产业的自动化程度
			传统产业排污费与 GDP 比
			产业聚集性
			万元 GDP 能耗
			电子商务交易额
		智慧产业	智慧产业固定资产投资额
			智慧产业 R&D 经费支出
			智慧产业占 GDP 比重
			智慧产业从业人员数
			智慧产业年发明专利申请总量

(续表)

属 性	9 个 一级指标	22 个 二级指标	99 个 三级指标
智能之花	智慧民生	教育	优质教育资源的均衡智能配置程度
			优质教育资源的覆盖程度
			幼儿园数量最低区域/每平方公里/每万人
			中小学数量最低区域/每平方公里/每万人
		社保	社会医保一卡通使用率
			居民社保覆盖率
			保险、福利、养老服务便捷程度
			保险、福利、养老服务的监督水平
		医疗	居民获得医疗服务的满意度
			药品的溯源性
			药品供应事故率
			服务质量监督的透明度
	智慧社会	交通	人均交通卡拥有数
			城市交通引导系统覆盖程度
			公交站牌电子化率
		居住	弱势群体租房服务的智能性
			弱势群体租房服务的透明性
			弱势群体人均居住面积
		社会服务	政府信息公开种类和条目数量
			政务门户网站网上办事范围和效率
			就业信息服务使用率
			空气质量监测与服务的信息发布频次
			地表水环境质量监测与服务的信息发布频次
			环境噪声监测与服务的信息发布频次
			污染源监控与服务的信息发布频次
			城市饮用水环境监测与服务的信息发布频次
		社会管理	土地资源监测、土地利用变化监测的智能化程度
			应急救援物流的智能化程度
			灾害预警系统的智能化程度
			应急指挥系统的智能化程度

(续表)

属 性	9 个 一级指标	22 个 二级指标	99 个 三级指标
智能之花	智慧社会	社会管理	社区管理的数字化程度
			金融诚信体系监管智慧化程度
			金融安全体系智慧化程度
			在线支付便捷性程度
			在线支付安全性程度
			食品的溯源性
			食品供应事故率
	智慧环境	环境宜居性	体育设施覆盖度和使用率
			文化娱乐设施覆盖度和使用率
			人均绿地面积
			城区绿化覆盖率
		污染指数	空气污染指数
			城镇生活污水处理率
			工业固体废物综合利用率
			三废综合利用产品产值
智能之干	智慧治理	智慧规划 体系	智慧城市发展规划纲要及实施方案的完整性和可行性。
			负责智慧城市创建的组织体系和执行机构完善比例
			保障智慧城市建设和运行的政策法规制定和落实的完善比例
			智慧城市建设经费预算和保障措施数
			智慧城市运行监督体系完善率
		城市功能 完善性	通勤换乘次数均值
			建筑智能化水平
			城市景观风貌智能化程度
			历史文化保护智能化程度
	智慧人才	人才结构	大专以上学历人口比重
			科技人员人口比重
			创意产业人口比重
			信息服务业从业人员占全市从业人员比重
	智慧平台	公共基础 数据库	地理信息数据库完整性、及时性
			地理信息数据库便捷性、安全性

(续表)

属 性	9 个 一级指标	22 个 二级指标	99 个 三级指标
智能之干	智慧平台	公共基础 数据库	人口基础数据库完整性、及时性
			人口基础数据库便捷性、安全性
			法人基础数据库完整性、及时性
			法人基础数据库便捷性、安全性
			宏观经济数据库完整性、及时性
			宏观经济数据库便捷性、安全性
		公共信息 平台	各级各类政务服务平台公共信息的统一性
			各级各类政务服务平台公共信息的数据共享受性
			各级各类政务服务平台的连接性与融合度
			政务服务平台的用户覆盖率
智慧之根	智慧基础 设施	网络基础 设施	无线网络覆盖率
			光纤接入覆盖率
			户均网络带宽
		感知基础 设施	城市道路传感器覆盖数量
			智能表具安装普及率
			路灯智能化自动化控制率
			空气质量监测 100 km ² 检测点数
			地表水环境空气质量监测 100 km ² 检测点数
			环境噪声监测 100 km ² 检测点数
			污染源监控 100 km ² 检测点数
			城市饮用水环境 100 km ² 检测点数
智慧之壤	自然禀赋	交通基础 设施	公路、铁路(地铁)单位面积里程
			每 10 km 沿岸港口数
			每 200 km 内机场数
		自然资源 禀赋	地理特征
			环境气候特征
		人文资源 禀赋	土地、矿产资源种类、储备
			历史古迹保护等级、数量
			文化遗产等级、数量

7.4 安全隐私

智慧城市充分利用信息化相关技术,通过监测、分析、整合以及智慧响应的方式,综合各职能部门,整合优化现有资源,提供更好的服务、绿色的环境、和谐的社会,保证城市可持续发展,为企业及大众建立一个优良的工作、生活和休闲环境。形象地说,智慧城市=全面透彻的感知+宽带泛在的互联+智能融合的应用+以人为本的创新,如图7-2所示。

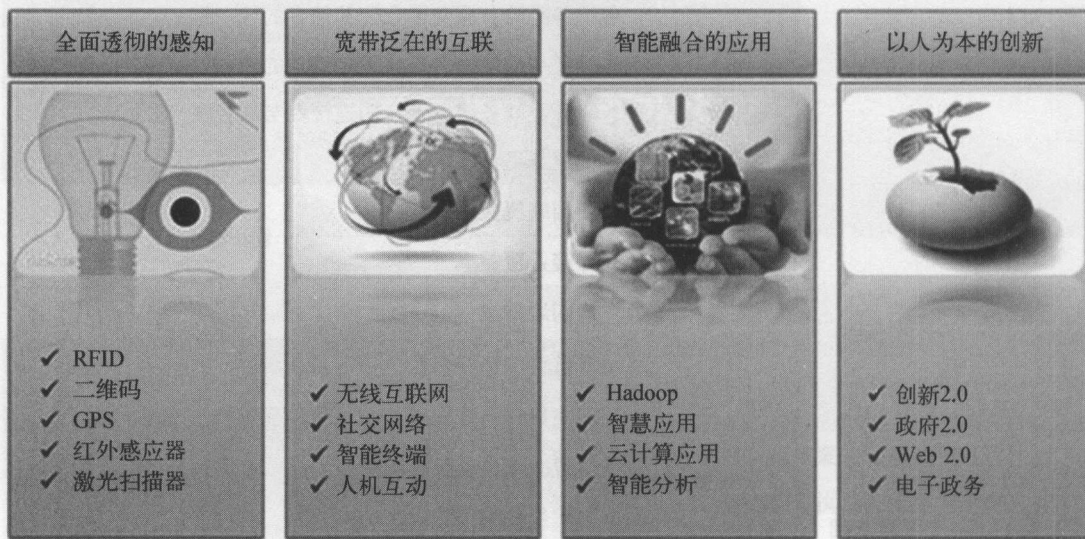


图7-2 智能城市的信息化构成

由于智慧城市的复杂性,使得智慧城市的信息安全问题更加复杂,智慧城市的信息安全保障是针对支撑性信息网络环境而言的,这包括系统的信息感知层、接入与传达层、应用于终端层、智能/智慧处理及协同层等,对于智慧城市典型的安全威胁包括:

1) 安全隐私鉴别与保密

RFID技术被广泛地用于物联网系统时,RFID标签被嵌入任何物品中,比如人们的日常生活用品中,而用品的拥有者不一定能觉察,从而导致用品的拥有者不受控制地被扫描、定位和追踪,带来隐私保护的问题,这不仅涉及技术层面,还将涉及法律和道德层面。

2) 智能感知节点的自身安全问题

即物联网机器/感知节点的本地安全问题。由于物联网的应用可以取代人来完成一些复杂、危险和机械的工作,所以物联网机器/感知节点多数部署在无人监控的场景中。那么攻击者就可以轻易地接触到这些设备,从而对它们造成破坏,甚至通过本地操作更换机器

的软硬件。

3) 物联网的业务安全

由于物联网节点无人值守,并且有可能是动态的,所以如何对物联网设备进行远程签约信息和业务信息配置就成了难题。另外,现有通信网络的安全架构都是从人与人之间的通信需求出发的,不一定适合以机器与机器之间的通信为需求的物联网。使用现有的网络安全机制会割裂物联网机器间的逻辑关系。

4) 更容易传播恶意代码

实现智慧城市的前提是互联,这包括了人与人、人与物、物与物等多个层面的互联互通,也包含了资源的融合与共享,如跨政府职能部门、跨运营商、跨金融企业等的资源互通与互访,而一旦某个环节受到了恶意代码的侵入,将使得威胁的蔓延会更快。

5) 超大规模的拒绝服务攻击

由于物联网中节点数量庞大,且以集群方式存在,因此在数据传播时,大量节点的数据传输需求会导致网络拥塞,这对于现有的基础网络结构带来了新的挑战;另外,由于大型物联网节点鉴别技术的复杂性,如果没有很好的检测机制,利用智慧城市的信息平台发动大规模甚至超大规模的攻击,也将会更加容易实现。

6) 云计算技术的应用带来的安全挑战

云计算技术是实现海量接入和并行计算的新型技术,也必然成为支撑智慧城市的典型信息技术手段,而云技术引入带来的安全威胁,是当前云业务发展乃至未来网络社会的最大挑战,这包括用户和服务商之间未建立起信任关系;云服务核心技术仍无法摆脱受制于人的境地,不掌握核心技术就是最大的安全隐患;我国云安全标准目前还处于空白状态,阻碍我国云服务快速发展;传统的安全技术在云虚拟化环境中,依然受到了更大的挑战。

7.4.1 物联网的隐私维护和信息安全

从技术角度看,物联网的基础设施建立在数据通信工具之上(主要是贴有电子标签即RFID的物品)。很多物联网是基于EPCglobal和GS1的电子产品代码(Electronic Product Code)技术而建立起来。物联网中的设备或物品被贴上存有唯一电子产品代码的电子标签。物联网可以为用户提供电子产品代码信息服务(Electronic Product Code Information Services, EPCIS)。RFID芯片不一定存储设备或物品的所有信息,部分信息可以存在分布于不同地理位置的服务器上,这些信息可以通过对象命名系统(Object Naming System, ONS)提供的链接地址提取出来。对象命名系统可以成为普适计算(Ubiquitous Computing)的基础架构,智能城市信息系统可以识别物联网上的智能设备和物品,并通过互联网获取它们的所有信息。对象命名系统的运作方式和域名系统(Domain Name System, DNS)类似,通过把电子产品代码EPC转化成类似DNS域名的格式并进一步转化成二进制地址,智能城市信息系统可以根据二进制地址找到物联网上的设备和物品。

以上提到的物联网基础设施技术对物联网的隐私和信息安全有重要意义。要达到对隐私和信息安全的保护,对智能城市信息系统提出了以下要求:

(1) 对攻击的耐受能力。系统需根据单个节点的运行状况及时自我调整,防止因单个节点受到攻击导致整个系统崩溃。

(2) 数据验证。作为一项原则,物联网设备地址和信息的提取必须进行验证,确保获得的数据没有在传输过程中遭到篡改。

(3) 访问控制。信息提供者必须能实现对数据的访问控制。

(4) 客户隐私。确保只有信息提供者能够根据对客户行为的观察推断出客户身份。简而言之,根据可观察到的行为推断客户身份应该是非常困难的。

智能城市信息系统应该充分考虑到以上的系统要求,从而对系统实施风险管理。现有一些技术可用于加强控制客户的隐私和信息安全。

(1) 虚拟专用网(Virtual Private Networks, VPN)是商业伙伴之间在公用网络上建立的封闭的专用网络,进行加密通信。在企业网络中有广泛应用。由于仅有商业伙伴可以访问 VPN,这确保了通信的私密性和完整性。但是 VPN 解决方案不利于实现动态的全面信息共享,因此对于 VPN 之外的第三方不具可操作性。

(2) 安全传输层协议(Transport Layer Security, TLS)基于全局的信任机制,也可以提高物联网通信的私密性和完整性。但是当每个 ONS 授权都需要 TLS 连接时,信息搜索效率就会降低。

(3) DNS 安全扩展(DNS Security Extensions, DNSSEC)使用公钥加密方法对记录签名,确保信息源记录的真实性和完整性。但是 DNSSEC 仅在所有的物联网组织都接受它的情况下才能确保 ONS 信息的真实性。

(4) 洋葱路由(Onion Routing)。通信数据先进行多层加密然后在由若干个被称为洋葱路由器组成的通信线路上被传送。每个洋葱路由器去掉一个加密层,以此得到下一条路由信息,然后将数据继续发往下一个洋葱路由器,不断重复,直到数据到达目的地。这就防止了那些知道数据发送端以及接收端的中间人窃得数据内容。但是这种方法增加了等待时间,因此降低了系统的性能。

(5) 个人信息检索(Private Information Retrieval, PIR)系统可以在使用电子产品代码信息服务(EPCIS)时隐藏哪个客户对哪些信息感兴趣。但是系统的扩展性(Scalability)和关键字管理(Key Management)都受到影响,同时也影响那些开放访问的服务,比如 ONS 的性能。

杨金翠和方滨兴(2011)提出以下的物联网信息安全模型,如图 7-3 所示。控制模块通过感应器件(Sensors)和通信模块建立联系。计算模块通过网络和通信模块建立联系。而信息-物理融合系统(Cyber-Physical System)将计算模块和控制模块融合起来。在计算模块和控制模块的融合中,需要对信息安全进行控制(例如防止黑客冒充成合法使用者操纵控制模块)。

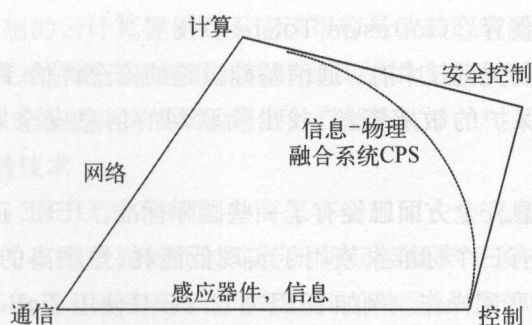


图 7-3 智能城市的物联网信息安全模型

杨金翠和方滨兴(2011) 还提出了物联网信息安全的框架(图 7-4)。

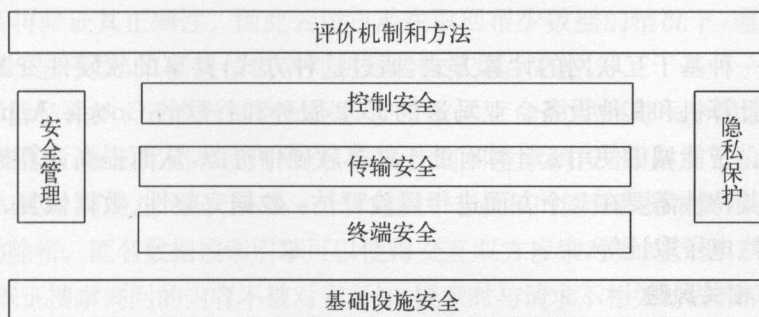


图 7-4 智能城市的物联网信息安全的框架

在基础设施安全方面的关键技术有：可用于多个身份标签的 RFID 安全基础架构、全闭环系统和微界面的设计、根据风险自动调节的资源分配机制、物联网的安全保障模型、物联网的信任机制模型、物联网的可信任传输方法。

在终端安全方面的关键技术有：RFID 和身份管理系统、识别低能耗的可信任的终端、标签信息的访问控制、标签信息的读操作验证、数据冗余、备份、验证和存储技术、降低交互次数的通信协议技术。

在传输安全方面的关键技术有：终端通信的加密和验证技术、传感器网络的反攻击和冗余通信渠道技术、低能耗、低复杂度和高安全的路由协议、少量数据链接的加密技术、物联网和互联网、3G 网络的集成。

在控制安全方面的关键技术有：控制失败的检测技术、传输冗余技术、冗余控制渠道技术、控制的冗余和生存期、验证和决定控制的真实性技术、防止污染的控制技术、获得控制的反馈和可用性技术。

在隐私保护方面的关键技术有：智能终端的反克隆技术、可信任的智能终端技术、智能终端的自毁技术、匿名通信技术、实时可靠的 Agent 技术、多来源的身份管理方法和隐私保护、安全验证技术、统一的用户管理技术。

在信息安全方面的关键技术有：密钥管理、加密算法管理、物联网阶段的协同运作的验

证和管理、容错技术、入侵容忍(Intrusion Tolerant)技术。

在评价机制和方法的关键技术有：通信基础设施的安全评估、网络安全政策的设计和评价、定位物联网的需保护的敏感资源、找出物联网的信息安全风险、物联网的脆弱性分析。

目前在物联网的信息安全方面已经有了一些国际标准。IEEE 正在对物理层和介质访问控制层进行标准化。有三个标准家族用于实现低能耗、短距离的机器到机器的(M2M, Machine-to-Machine)物联网操作。例如 IEEE 802. 15. 4(使用 ZigBee)、IEEE 802. 15. 1(使用蓝牙)和 IEEE 802. 15. 11(使用 Wi-Fi)。

7.4.2 云计算的隐私维护和信息安全

云计算是一种基于互联网的计算方式,通过这种方式,共享的软硬件资源和信息可以按需求提供给计算机和其他设备。亚马逊的 EC2 服务和谷歌的 Google App Engine 都是云计算的例子。智能城市使用云计算有助于共享软硬件资源,从而提高计算机资源的利用率。云计算因其特性需要在几个方面进行风险评估:数据完整性、数据恢复、隐私、法律问题如合规、审计、电子取证等。

1) 云计算相关风险

(1) 特许用户对数据的访问。敏感数据放在云计算提供商的服务器和数据库上,可能带来信息安全和隐私泄露的风险。这是因为云计算提供商有可能忽略掉“物理、逻辑和人员的控制”,而这些控制是在不使用云计算而自主管理数据时需要做的。所以需要详细询问云计算提供商如何聘用可以接触到敏感数据的特许用户(如数据库管理员)以及如何对这些特许用户的数据访问权限进行控制。

(2) 合规(Regulatory Compliance)。使用云计算服务的用户最终要为信息安全和数据完整性负责,即使数据放在云计算服务商那里。需要让服务商接受一系列的外部审计和信息安全认证,以确保信息安全和隐私得到保护。

(3) 数据存放地点。云计算服务商可能将数据存储在不同的城市甚至另外一个国家。某些国家的法律对隐私和信息安全的要求可能比较低,这时需要确保云计算服务商提供符合云计算使用者的隐私和信息安全要求的服务。

(4) 数据隔离。云上的数据是在一个多用户共享的环境下进行存储的,需要保证其他用户不能接触到自己的数据。加密是有效的方式,但要确保不会发生加密失败事故,导致数据无法读出和使用。

(5) 数据恢复。需要了解云计算服务商如何在发生灾难事故的时候对数据进行恢复。如果云计算服务商不能备份数据,那么智能信息系统就有很大的脆弱性。

(6) 支持调查。应该通过合约形式要求云计算服务商对事后调查提供支持,比如提供登录系统的日记数据,以便在发生信息安全事故的情况下展开调查。

(7) 长期生存。理想的云计算提供商不应该很容易就破产或者被其他公司兼并。但是如果发生这种情况,也要确保数据能够顺利转移到其他供应商那里。所以需要和云计算提供商确认数据的存储格式能够转换成其他通用的存储格式。

2) 云计算安全关键技术

(1) 可信访问控制: 由于无法信赖服务商忠实实施用户定义的访问控制策略,所以在云计算模式下,研究者关心的是如何通过非传统访问控制类手段实施数据对象的访问控制。其中得到关注最多的是基于密码学方法实现访问控制。基于密码类方案面临的一个重要问题是权限撤销,一个基本方案是为密钥设置失效时间,每隔一定时间,用户从认证中心更新私钥。

(2) 数据存在与可使用性证明: 由于大规模数据所导致的巨大通信代价,用户不可能将数据下载后再验证其正确性。因此云用户需在取回很少数据的情况下,通过某种知识证明协议或概率分析手段,以高置信概率判断远端数据是否完整。

(3) 数据隐私保护: 云中数据隐私保护涉及数据生命周期的每一个阶段,包括云中数据生成与计算阶段、数据存储和使用阶段、数据查询和搜索阶段、数据发布阶段等。例如隐私保护系统 Airavat 可防止 Map Reduce 计算过程中非授权的隐私数据泄露出去,并支持对计算结果的自动除密。匿名数据搜索引擎可以使得交互双方搜索对方的数据,获取自己所需要的部分,同时保证搜索询问的内容不被对方所知,搜索时与请求不相关的内容不会被获取。

(4) 虚拟安全技术: 使用虚拟技术的云计算平台上的云架构提供者必须向其客户提供安全性和隔离保证,例如在 Grid 环境下隔离执行机。

(5) 云资源访问控制: 在云计算环境中,各个云应用属于不同的安全管理域,每个安全管理域都管理着本地的资源和用户。当用户跨域访问资源时,需要在域边界设置认证服务,对访问共享资源的用户进行统一的身份认证管理。

(6) 可信云计算: 以可信赖方式提供云服务已成为云安全研究领域的一大热点。服务商可以向其用户提供一个密闭的箱式执行环境,保证客户虚拟机运行的机密性。另外,服务商还允许用户在启动虚拟机前检验服务商的服务是否安全。可信计算技术提供了可信的软件和硬件以及证明自身行为可信的机制,可以用来解决外包数据的机密性和完整性问题。

◇ 参 考 文 献 ◇

[1] 佚名. 美国政府大数据网站 data.gov 即将开源. <http://www.topoint.com.cn/html/article/2012/>

- 12/335787.html.
- [2] Yuxin. 大数据的积极拥趸者——英国. <http://www.jifang360.com/news/2014124/n997855993.html>.
 - [3] 佚名. 法国大数据: 智慧城市中的大数据. <http://datacenter.ctocio.com.cn/51/12894051.shtml>.
 - [4] 中云网. 日本: 用大数据创建最尖端 IT 国家. <http://www.china-cloud.com/dashujuzhongguo/disangqi/2014/0114/22689.html>.
 - [5] 佚名. 纽约市利用大数据防火. http://news.xinhuanet.com/info/2014-02/12/c_133108441.htm.
 - [6] Dink. 上海通信管理局: 截至 2013 年 10 月上海市 3G 移动电话用户数 1 074 万户. <http://www.199it.com/archives/181004.html>.
 - [7] 佚名. 中国电信三大举措推动 2013 宽带发展. <http://www.miit.gov.cn/n11293472/n11293877/n15432927/n15433065/15546792.html>.
 - [8] 佚名. 中国联通三大举措推动 2013 宽带发展. <http://www.miit.gov.cn/n11293472/n11293877/n15432927/n15433065/15546925.html>.
 - [9] 佚名. 日本“大数据”有望年均创收 7 077 万亿日元. http://news.xinhuanet.com/newmedia/2013-07/17/c_125020131.htm.
 - [10] 天虹. IDC: 2017 年大数据市场规模将达 324 亿美元. http://news.ccidnet.com/art/1032/20131221/5296989_1.html.
 - [11] 无忌. 美国风投 2013 年五大最爱消费电子成为新宠. <http://tech.qq.com/a/20140212/019749.htm>.
 - [12] 曹华. 国开行今后三年投融资 800 亿元支持智慧城市建设. <http://finance.people.com.cn/bank/n/2013/0114/c202331-20194680.html>.
 - [13] 张茜翼. 中国智慧城市建设风生水起. <http://money.163.com/13/1017/21/9BDVNRE500254TI5.html>.
 - [14] 郑玲微. 智慧城市亟需标准破题. <http://www.chinaeag.gov.cn/show-5244.html>.
 - [15] Banker. 盘点 2013 年中国各行业数据量. <http://bbs.middleware123.com/thread-24066-1-1.html>.
 - [16] 佚名. 大数据业务的商业模式探讨. <http://www.e-gov.org.cn/xinxihua/news004/201307/142489.html>.
 - [17] 佚名. 大数据决定未来业务模式和发展方向. <http://info.edu.hc360.com/2013/02/280835557647.shtml>.
 - [18] 顾德道, 乔雯. 我国智慧城市评价指标体系的构建研究[J]. 未来与发展, 2012(10): 79-83.
 - [19] 张本伟. 加快建设智慧城市 提升新型城镇化水平[J]. 商情, 2013.

后记

本书的出版，首先要感谢我的妻子。她不仅是我生活中的伴侣，更是我精神上的支柱。在本书的写作过程中，她给予了我无私的支持和鼓励。其次，我要感谢我的家人，特别是我的父母，他们一直是我前进的动力。最后，我要感谢我的朋友们，他们在我遇到困难时给予了我及时的帮助。本书的出版，离不开许多人的帮助，在此一并致谢。本书的出版，也离不开许多人的帮助，在此一并致谢。本书的出版，也离不开许多人的帮助，在此一并致谢。

本书的出版，首先要感谢我的妻子。她不仅是我生活中的伴侣，更是我精神上的支柱。在本书的写作过程中，她给予了我无私的支持和鼓励。其次，我要感谢我的家人，特别是我的父母，他们一直是我前进的动力。最后，我要感谢我的朋友们，他们在我遇到困难时给予了我及时的帮助。本书的出版，离不开许多人的帮助，在此一并致谢。本书的出版，也离不开许多人的帮助，在此一并致谢。本书的出版，也离不开许多人的帮助，在此一并致谢。

从智慧城市建设的角度,如果我们把智慧城市的建设比喻成一棵树的话,这其中树叶是末梢,离不开物联网、移动计算等技术,强调的是动态感知;树枝、树干的成长离不开云计算、SDN 等关键技术,强调的是互联互通和协同共享;而这其中根基是大数据等技术,强调的是智能化、分析处理和可视化。可见大数据在整个智慧城市建设中的重要地位和作用。

2013 年 5 月,奥巴马发布了新的“政府信息实现开放共享”法规。2013 年 8 月,澳大利亚政府信息管理办公室(AGIMO)发布了公共服务大数据战略。2014 年 2 月,爱尔兰政府宣布建成国家大数据分析研究中心——Insight。大数据的使用已经成为一个国家各领域提高生产力、创新能力以及竞争力的关键要素,大数据已经成为国家战略。现今大数据已成为 21 世纪的战略资源,成为发达国家战略布局和跨国企业资金投入的重点。

虽然我国在智慧城市建设和大数据的开发和利用方面取得了一定的成绩,但是在智慧城市的现今阶段,我们也要清醒地看到大数据的开发和利用所面临的困难:

- 政府各个部门之间的数据共享程度低
- 政府数据开放刚刚起步
- 大数据生态环境中各个角色的权利义务不清
- 数据资产所有权的界定和保护
- 数据利用中的安全和隐私保护问题
- 处理大数据能力的技术和平台
- 大数据相关的人才匮乏

面向未来,对于在智慧城市的建设过程中如何更好地利用大数据,需要我们进一步的思考、解决和实践:

(1) 数据资产的共享、整合、协同、盘存、购买、活化:做好顶层设计、了解需求,从组织架构、体制和机制上予以保障;要摸清家底,到底我们有哪些数据资源,有哪些不同的类别,是内部共享、对外协同服务、社会第三方公开,还是外部采购;我们要强化已有数据资源的活化,保持实时性和及时性;支持和投入对大数据相关共性技术的研究与攻关。

(2) 建立政府数据开放的机制:参考国外政府大数据开放所遵循的原则,完整性、原始性、及时性、可获取性、可处理性、非歧视性、非专有性、非许可性等;鼓励政府部门共享和开放,纳入对政府相关部门的绩效考核;建立数据质量评估机制,纳入对政府相关部门的绩效考核;鼓励第三方挖掘政府信息资源的积极性,出台相关的奖励措施。

(3) 加强对数据隐私、安全、知识产权、法律法规、标准规范的研究与制定:建立相关标准规范、法律法规;研究个人数据保护的法律瓶颈;明确拥有者、使用者、第三方社会资源、最终用户等各方的责任、权利和义务;加强对数据主权研究;加强对数据跨境流动监管;加强数据风险分级管理。

(4) 加强大数据人才(分析、管理、技术)培养与培育:尤其注重培育跨界复合型人才,既要熟悉政府业务,又要熟悉大数据技术;加强对政府信息化和管理决策部门、对大数据技术人才的相关培训,通过交流沟通寻找技术与业务的结合点。

(5) 推动技术、产品与服务的加速融合与创新:鼓励多种形式的服务模式的探索和创新;鼓励政府部门购买大数据相关的产品和服务;鼓励社会第三方基于大数据的运营模式探索和创新;支持大数据公共服务平台、专业大数据公共平台、产学研联盟以及相关基础设施的建设。

大数据时代真的已经来临,智慧城市的建设更加离不开大数据相关技术和产品的应用、服务和管理模式的创新、融合业务和运营模式的探索,我们坚信,智慧城市的美好蓝图一定会早日实现。大数据不仅是引领新一轮城市经济社会发展的重要驱动因素,一定意义上也将是人类社会发展进程中的划时代变革。

索引

城市群, 36

大数据, 52

大数据采集, 78

大数据处理, 78

大数据存储, 78

大数据决策分析, 78

功能型城市, 35

集市型城市, 35

指标体系, 185

智慧城市, 5

综合型城市, 36



策划编辑 张 晨 魏 玲 玲 陈 立
责任编辑 姚晨晖
装帧设计 房惠平

大数据技术与应用

- 汇计划在行动
- 大数据测评
- 数据密集型计算和模型
- 城市发展的数据逻辑
- 智慧城市大数据
- 金融大数据
- 城市交通大数据
- 医疗大数据



上海科学技术出版社
www.sstp.cn

上架建议
大数据·计算机技术·城市规划

ISBN 978-7-5478-2290-6



9 787547 822906 >

定价：52.00元
易文网 www.ewen.co